

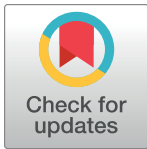
RESEARCH ARTICLE

A simplified modelling framework facilitates more complex representations of plant circadian clocks

Mathias Foo¹, Declan G. Bates², Ozgur E. Akman^{3*}

1 School of Mechanical, Aerospace and Automotive Engineering, Coventry University, Coventry, United Kingdom, **2** Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, United Kingdom, **3** College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom

* O.E.Akman@exeter.ac.uk



Abstract

The circadian clock orchestrates biological processes so that they occur at specific times of the day, thereby facilitating adaptation to diurnal and seasonal environmental changes. In plants, mathematical modelling has been comprehensively integrated with experimental studies to gain a better mechanistic understanding of the complex genetic regulatory network comprising the clock. However, with an increasing number of circadian genes being discovered, there is a pressing need for methods facilitating the expansion of computational models to incorporate these newly-discovered components. Conventionally, plant clock models have comprised differential equation systems based on Michaelis-Menten kinetics. However, the difficulties associated with modifying interactions using this approach—and the concomitant problem of robustly identifying regulation types—has contributed to a complexity bottleneck, with quantitative fits to experimental data rapidly becoming computationally intractable for models possessing more than ~ 50 parameters. Here, we address these issues by constructing the first plant clock models based on the S-System formalism originally developed by Savageau for analysing biochemical networks. We show that despite its relative simplicity, this approach yields clock models with comparable accuracy to the conventional Michaelis-Menten formalism. The S-System formulation also confers several key advantages in terms of model construction and expansion. In particular, it simplifies the inclusion of new interactions, whilst also facilitating the modification of regulation types, thereby making it well-suited to network inference. Furthermore, S-System models mitigate the issue of parameter identifiability. Finally, by applying linear systems theory to the models considered, we provide some justification for the increased use of aggregated protein equations in recent plant clock modelling, replacing the separate cytoplasmic/nuclear protein compartments that were characteristic of the earlier models. We conclude that as well as providing a simplified framework for model development, the S-System formalism also possesses significant potential as a robust modelling method for designing synthetic gene circuits.

OPEN ACCESS

Citation: Foo M, Bates DG, Akman OE (2020) A simplified modelling framework facilitates more complex representations of plant circadian clocks. PLoS Comput Biol 16(3): e1007671. <https://doi.org/10.1371/journal.pcbi.1007671>

Editor: Pedro Mendes, University of Connecticut School of Medicine, UNITED STATES

Received: March 8, 2019

Accepted: January 21, 2020

Published: March 16, 2020

Copyright: © 2020 Foo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All MATLAB files used to generate the results presented in the study are available from <https://github.com/mathiasfoo/essystemplantcircadian>.

Funding: MF gratefully acknowledges the financial support of The Royal Society via research grant RGS/R2/180195. DGB gratefully acknowledges the financial support of the EPSRC and BBSRC via research grant BB/M017982/1. OEA wishes to acknowledge the financial support of the EPSRC (research grants EP/N017846/1 and EP/N014391/1). The funders had no role in the study design,

data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Mathematical models have been widely employed as a complement to experimental work in elucidating the underlying mechanistic behaviour of the plant circadian clock. In this study, we investigate the use of a simplified modelling strategy, the S-System framework, to reduce the computational complexity of such models. We test the efficacy of our approach by constructing S-System versions of five established plant clock models, which we fit to synthetic and experimental gene expression data. We demonstrate that the S-System formulation can generate fits to expression data that are very similar to those of its more complex counterparts, yielding predictions with better accuracy in some cases. Our work reveals three principal benefits of the S-System formalism in modelling gene regulatory networks (GRNs) characterised by entrainable oscillations, typified by the circadian clock. Firstly, it enables new interactions to be incorporated in a very easy manner. Secondly, it facilitates the inference of unknown regulation types. Thirdly, it mitigates the parameter identifiability problem, thereby making it a suitable modelling method for designing GRN control circuits for synthetic biology applications. Taken together, our findings identify the S-System framework as a promising alternative approach for computational modelling of oscillatory GRN networks.

Introduction

Circadian clock networks

Most living organisms possess innate molecular clock machineries that govern their daily activity [1]. These machineries, known as circadian clocks, are responsible for the generation of endogenous oscillations in gene expression with a period close to 24 hours. Circadian oscillations enable the anticipation of diurnal environmental changes and the coordination of biological processes to occur at the optimal time of day. Some important biological functions that are circadian regulated include the mammalian sleep/wake cycle, fungal spore formation and plant leaf movement (see *e.g.* [2–4]). Moreover, interruption of the circadian rhythm can lead to a number of pathophysiological conditions, including poor metabolism, psychiatric disorders and immune system dysfunction (see *e.g.* [5–8]). At the molecular level, the underlying core mechanism of the circadian rhythm is generated via interlocking feedback loops between regulatory genes. This discovery was first made in the fruitfly and its significance is evident in the award of the 2017 Nobel Prize in Physiology or Medicine to the pioneers of molecular circadian systems research [9, 10].

Modelling the plant circadian clock

The discovery of multiple plant circadian genes has revealed the complexity of the underlying gene regulatory network (GRN), driving the use of mathematical models to help unravel the mechanisms controlling circadian timekeeping, as evident through the active development of clock models in the higher plant *Arabidopsis thaliana*. Since the construction of the first *Arabidopsis* model in 2005 by Locke *et al.* [11], which comprised only two key clock genes, the models have expanded considerably in size and complexity, with frequent updates to incorporate more interactions as new experimental data became available. From a modeller's perspective, updating a model to include new interactions is greatly facilitated if the model structure has low inherent complexity—*i.e.* an efficient parametrisation and minimal nonlinear terms, in

the case of differential equation models. Overly complex models incur a high computational cost in terms of parameter optimisation [12, 13], thereby limiting their predictive capacity. They also tend to cloud the core behaviour of the system of interest, particularly the identification of possible network motifs. In view of this, many studies have focused their attention on reducing model complexity, including some notable examples that specifically addressed the plant clock.

Reducing model complexity in plant clock models

In [14], instead of considering all circadian genes individually, the authors grouped several key circadian genes together and analysed the behaviour of the resulting reduced set of differential equations. In [15], the authors reduced the complexity of the model by identifying the *kernel* of the GRN—those genes that are solely accountable for generating circadian rhythms with behaviour similar to that of the wild-type.

In [12], the authors focused on reproducing key circadian characteristics (*e.g.* entrainment and photoperiodism) through the use of continuous-time Boolean models. The Boolean framework, in which genes are assumed to be either ‘OFF’ (0) if expression is below some threshold or ‘ON’ (1) if expression is above the threshold, yielded the following two key reductions in complexity: (i) all the parameters governing the time taken for the production of a transcription factor (TF) and its effect on a downstream gene (*e.g.* transcription rates, translation rates, degradation rates *etc.*) are telescoped into a single delay parameter; (ii) the complex nonlinear functions governing the expression of each gene are replaced by a constrained set of Boolean functions (logic gates), meaning that all the architectures consistent with a given circuit diagram can be systematically explored. In related work, distributed delays were used to represent TF production pathways in three established clock models. This approach replaces the set of parameters governing the delays in each pathway with a pair of parameters that control the mean and variance of the delay distribution (assumed to be a gamma function), leading to simplified differential equation models with markedly reduced parameterisations [16].

While those aforementioned approaches do dramatically reduce model complexity, they can come at the expense of either reducing the ability of the model to match quantitative behaviour (*e.g.* an inability to simulate amplitude modulation in the case of the Boolean models) or of favouring fits to specific biological phenotypes.

S-Systems—A simplified framework for modelling the plant circadian system

The task of developing mathematical models of reduced complexity that still preserve accuracy presents a significant challenge to modellers. In the context of plant circadian systems biology, the network of identified clock genes is set to increase in size, given the extensive research in this area. A model structure is therefore required that facilitates modification and updating as further progress is made experimentally. The majority of plant clock models developed thus far have been sets of nonlinear differential equations based on Michaelis-Menten kinetics (see *e.g.* [11, 17–19]). However, modifying models based on this formalism to incorporate additional interactions (or revise existing ones) in the light of new experimental data can significantly increase computational complexity. This is predominately due to the inherent structure of Michaelis-Menten models, for which different regulation types (*i.e.* transcriptional activation and inhibition) are modelled with different nonlinear functions. The effect of this functional heterogeneity is further amplified when multiple transcriptional regulators are combined. Accordingly, if larger systems (*e.g.* of the order of 100 parameters or more) are to

become amenable to quantitative modelling, alternative model formulations that mitigate this issue are required.

Here, we develop a minimal framework for modelling the plant circadian system using ordinary differential equations (ODEs), focusing mainly on simplifying the nonlinear functions governing gene expression. Our framework assumes a simple, homogenous model structure based on an extension of the S-System formalism originally developed by Savageau [20] to model biochemical systems, with specific modifications that enable the interactions between plant circadian genes and the photic environment to be represented. We note that although S-Systems have been used to model gene regulatory networks previously (see *e.g.* [21, 22]), circadian clocks have not been modelled with this formalism to date.

We investigate the capacity of our framework to quantitatively reproduce circadian dynamics by constructing modified S-System versions of a suite of established plant clock models. By fitting these S-System formulations to synthetic and experimental expression timeseries, and validating our fits in each case against hold-out data, we demonstrate that our formalism yields models with comparable predictive power to their Michaelis-Menten based counterparts. In addition, by employing a frequency response analysis technique from linear systems theory, we provide a mechanistic understanding of the progressive simplification of protein production pathways adopted by more recent models of the plant clock.

Finally, we highlight the advantages conferred by the S-System framework in modelling GRNs and synthetic GRN controllers, and also discuss possible further extensions to the framework to facilitate model construction and increase prediction accuracy.

Materials and methods

S-System modelling

The S-System modelling framework arose from biochemical system theory (see *e.g.* [20]), with the initial purpose of describing metabolic pathways. The original S-System model introduced in [20] can be written in the form

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{i,j}}, \quad 1 \leq i \leq n, \quad (1)$$

where the dependent variables $\{X_1, \dots, X_n\}$ represent the biochemical species of interest and the independent variables $\{X_{n+1}, \dots, X_{n+m}\}$ represent forcing terms. For each X_i , α_i represents the production rate constant, and the $g_{i,j}$ s are the exponents associated with production processes, whilst β_i denotes the degradation rate constant, and the $h_{i,j}$ s are the exponents associated with degradation processes.

Over the course of its development, the S-System framework has been used as an alternative approach for modelling a broad range of biological processes (see [23, 24] and references therein). These include signal transduction [25], metabolism [26] and enzyme kinetics [27–29]. In [30], the authors compared the validity of S-System and Michaelis-Menten models of enzyme-catalysed reactions in a rigorous manner, demonstrating that the two models had similar accuracy over the same concentration ranges. The accuracy of these two formulations has also been compared when modelling fully developed pathway systems—*e.g.* the fermentation pathway in yeast [31], purine metabolism in humans [32] and sphingolipid metabolism in yeast [33]—where both formulations have demonstrated comparable performance. In [34], the authors compared different kinetic models for the flowering time GRN in *Arabidopsis*, finding that S-System and Michaelis-Menten formulations possessed similar predictive capacity. The design principles predicted by the S-System approach for gene regulation in [35, 36] and protein

modification in [37] have subsequently been verified in numerous experiments. The promising results of these studies prompted us to consider utilising S-Systems to model the plant clock.

Extending the S-System modelling framework to circadian clocks

The generation of circadian rhythms in plants (and other organisms) is primarily governed by three mechanisms: (i) transcription—the process in which one or more TFs can bind to the specific promoter region of a gene to regulate the conversion of DNA into RNA; (ii) translation—the process in which protein is created by ribosomes in the cytoplasm following RNA transcription, prior to moving into the nucleus to control transcription; and (iii) protein modification—processes such as protein stabilisation and/or degradation that mediate the efficacy of a TF [38].

In order to model these mechanisms, as represented in the clock circuits of interest, some modifications to Eq (1) are required. Firstly, we introduce an extra term that represents the contribution from external light inputs. Secondly, to account for (i) complex TF regulation of gene expression, and (ii) protein modification, we introduce suitable summation operations to the first and second terms of Eq (1). With that, the modified S-System model can be written as below:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n_i^p} \left(\sum_{k=1}^n b_{i,j,k} X_k \right)^{g_{i,j}} - \sum_{j=1}^{n_i^D} \beta_{i,j} X_i \left(\prod_{k=1}^n X_k^{h_{i,j,k}} \right) + \sum_{j=1}^{n_i^L} \gamma_{i,j} U_{ij}, \quad 1 \leq i \leq n. \quad (2)$$

For a given i , X_i is the expression level of the i th clock gene/protein species, and each $U_{ij} = U_{ij}(X_1, \dots, X_n, L_I(t))$ represents the effect on X_i of a process regulated by the external light signal $L_I(t)$. Note that the explicit dependence of U_{ij} on $\{X_1, \dots, X_n\}$ reflects the fact that our target models include the effect of light-regulated protein complexes on gene/protein expression, in addition to the effect of direct light regulation. For the clock models considered in this study, we assume that U_{ij} is a low-order polynomial function of its arguments. Furthermore, given the limited understanding and experimental evidence regarding the precise effect of light on many circadian genes and proteins (in terms of dimerisation *etc.*), no exponent is associated with U_{ij} .

For all models considered here, $L_I(t)$ is assumed to be a periodic square wave with minimum and maximum values of 0 and 1 respectively, and $t = 0$ is taken to correspond to dawn, meaning that $L_I(t)$ is given by

$$L_I(t) = \begin{cases} 1 & \text{if } 0 \leq t \bmod 24 < P; \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where P is the photoperiod. Accordingly, $P = 0$ and $P = 24$ correspond to constant dark (DD) and constant light (LL), respectively, while setting $P = 12$ simulates a symmetric light-dark cycle, *i.e.* alternating 12 hour periods of light and dark (12L:12D).

We further remark that in Eq (2), n_i^p , n_i^D and n_i^L denote the number of processes involved in the production, degradation and light regulation of X_i , respectively. As in the original S-System formulation of Eq (1), α_i represents the production rate constant of X_i , and the $g_{i,j}$ s are the exponents associated with production. The $b_{i,j,k}$ s are Boolean variables, $b_{i,j,k} \in \{0, 1\}$, which determine the species contributing to each particular production process. The degradation of X_i is determined by the rate constants of each contributing process $\{\beta_{i,j} : 1 \leq j \leq n_i^D\}$ together with the associated exponents $\{h_{i,j,k} : 1 \leq j \leq n_i^D, 1 \leq k \leq n\}$. The $\beta_{i,j}$ s can take both positive and negative values, with the former representing degradation and the latter representing stabilisation. Finally, the $\gamma_{i,j}$ s determine the strength of the light-regulated processes affecting expression of X_i . We note that the original S-System formulation can be recovered from Eq (2) by setting $n_i^p = n$, $b_{i,j,k} = \delta_{j,k}$, $n_i^D = n$, $\beta_{i,j} = \beta_i \delta_{i,j}$, $h_{i,j,k} = h_{i,k} - \delta_{i,k}$ and $n_i^L = 0$, where $\delta_{i,j}$, $\delta_{j,k}$ and $\delta_{i,k}$ denote the

Kronecker delta in each case. Moreover, the original S-System model (cf. Eq (1)) is itself a special case of the Generalised Mass Action (GMA) model [39, 40]. This has the general form

$$\frac{dX_i}{dt} = \sum_{k=1}^{T_i} \pm \gamma_{i,k} \prod_{j=1}^{n+m} X_j^{f_{i,j,k}}, \quad 1 \leq i \leq n, \quad (4)$$

where $\{X_1, \dots, X_n\}$ and $\{X_{n+1}, \dots, X_{n+m}\}$ again denote the dependent and independent variables, respectively. However, our extended S-System formulation (cf. Eq (2)) cannot, in general, be expressed in this form.

Having introduced our modified S-System modelling framework (hereafter termed the *extended S-System formalism*), we discuss in more detail below how it can be employed to describe the transcription, translation and protein modification mechanisms that are characteristic of our target clock models.

Extended S-System modelling of transcription. In general, there are two main types of transcriptional regulation: transcriptional activation, which increases gene transcription, and transcriptional inhibition, which decreases gene transcription. Conventionally, the mechanisms are modelled using a combination of Michaelis-Menten and Hill-type functions. For a gene G that is regulated by a single transcriptional regulator P , transcriptional activation is often modelled with an equation of the form

$$\frac{dG}{dt} = \frac{aP^n}{K^n + P^n} - bG, \quad (5)$$

while for transcriptional inhibition, the model structure is often given by

$$\frac{dG}{dt} = \frac{aK^n}{K^n + P^n} - bG. \quad (6)$$

In both equations, K is the Michaelis-Menten kinetic constant (threshold for activation/inhibition), a and b respectively parameterise the transcription and degradation rate constants and n represents the Hill coefficient (degree of binding cooperativity).

In the extended S-System formalism, both transcriptional activation and transcriptional inhibition, combined with the linear degradation used in Eqs (5) and (6), can be represented with the single model structure

$$\frac{dX_{i_G}}{dt} = \alpha_{i_G} X_{j_P}^{g_{i_G j_P}} - \beta_{i_G} X_{i_G}, \quad (7)$$

where $X_{i_G} = G$ and $X_{j_P} = P$ ($i_G \neq j_P$). In this formulation, $g_{i_G j_P} > 0$ models activation and $g_{i_G j_P} < 0$ models inhibition. As in the original S-System formulation, α_{i_G} represents the production rate constant of X_{i_G} (i.e. the transcription rate), $g_{i_G j_P}$ is the exponent associated with activation/inhibition, whilst β_{i_G} denotes the degradation rate constant. It should be noted that this model of transcriptional regulation holds a key advantage over the standard Hill-type approach of Eqs (5) and (6) when it comes to network inference, i.e. when one is interested in identifying the type of regulation at each node of the GRN from the available experimental data: the regulation type can simply be inferred from the sign of the fitted value of $g_{i_G j_P}$, thereby avoiding the use of more complex nonlinear terms capable of smoothly interpolating between activation and inhibition [41].

We further note that when a gene is regulated by multiple TFs, modellers are required to select whether the TF interactions should be represented using the continuous analog of a multi-input OR logic gate (in which the terms modelling the effect of each TF on gene

expression are summed together), the analog of a multi-input AND logic gate (in which the terms are multiplied together), or a combination thereof [12, 41, 42]. In plant clock modelling, inhibitors tend to be combined with other regulators using AND gates, whilst activators are combined using OR gates (e.g. [11, 43–45]). By default, the standard S-System model only implements the multi-input AND logic gate (we note that other logic gates can of course still be approximated with this approach, depending on the range of the input variables around the appropriate nominal operating point). Our extended S-System formulation, however, enables a broader set of multi-input logic gates to be natively represented, and in particular, the gates used in the plant clock models of interest.

Extended S-System modelling of translation. A standard approach to modelling translation in circadian clock models is to explicitly represent the shuttling of the translated protein between the cytoplasm (where translation occurs) and the nucleus (where the protein can regulate transcription) [38]. This shuttling mechanism, taken from [11, 43], is often described using the following pair of differential equations:

$$\begin{aligned}\frac{dP_C}{dt} &= aG - r_C P_C + r_N P_N - b_C P_C, \\ \frac{dP_N}{dt} &= r_C P_C - r_N P_N - b_N P_N.\end{aligned}\quad (8)$$

In the above, G is the gene, whilst P_C and P_N denote cytoplasmic and nuclear protein, respectively. a is the translation rate, r_N and r_C are the shuttling rates, and b_C and b_N denote the degradation rate constant of each protein form. This model assumes that the translated protein is not subjected to further activities such as complex formation, protein stabilisation and/or protein degradation.

Protein shuttling can be represented in the extended S-System formalism as

$$\begin{aligned}\frac{dX_{i_{P_C}}}{dt} &= \alpha_{i_{P_C}} X_{j_G}^{g_{i_{P_C}, j_G}} - \beta_{i_{P_C}} X_{i_{P_C}}, \\ \frac{dX_{i_{P_N}}}{dt} &= \alpha_{i_{P_N}} X_{i_{P_C}}^{g_{i_{P_N}, i_{P_C}}} - \beta_{i_{P_N}} X_{i_{P_N}},\end{aligned}\quad (9)$$

where $X_{j_G} = G$, $X_{i_{P_C}} = P_C$ and $X_{i_{P_N}} = P_N$ (with all three indices $\{j_G, i_{P_C}, i_{P_N}\}$ distinct). In Eq (9), $\alpha_{i_{P_C}}$ represents the production rate constant of $X_{i_{P_C}}$ (i.e. the translation rate) and $\{\beta_{i_{P_C}}, \beta_{i_{P_N}}\}$ are the protein degradation rate constants. For fixed values of the latter, the rate of protein shuttling between cytoplasm and nucleus is determined by the production rate constant $\alpha_{i_{P_N}}$ of $X_{i_{P_N}}$, together with the exponents $g_{i_{P_C}, j_G}$ and $g_{i_{P_N}, i_{P_C}}$.

Many of the earlier plant clock models (e.g. [11, 43]) used protein shuttling to promote oscillatory behaviour [38]. However, due to the limited knowledge available regarding time-dependent expression patterns of cytoplasmic and nuclear proteins in plant clock GRNs, it has now become common practice to consider the aggregated effect of these two proteins instead, representing them with a single ODE of the form below:

$$\frac{dP}{dt} = aG - bP. \quad (10)$$

In this simple model, P and G represent protein and gene respectively, whilst a is the translation rate and b is the degradation rate constant. Although this is already in our extended

S-System form, its most general representation is

$$\frac{dX_{i_p}}{dt} = \alpha_{i_p} X_{j_G}^{g_{i_p,j_G}} - \beta_{i_p} X_{i_p}, \quad (11)$$

where $X_{i_p} = P$, $X_{j_G} = G$ ($i_p \neq j_G$), and α_{i_p} and β_{i_p} are the translation and degradation rate constants, respectively. The exponent g_{i_p,j_G} in Eq (11) can model both linear and nonlinear dependence of protein production on mRNA dependence (*e.g.* setting $g_{i_p,j_G} = 1$ recovers Eq (10)).

The aggregated protein model of Eq (10) has been used in more recent plant clock models (*e.g.* see [44, 45]). Moreover, as we show later, linear systems theory (frequency response analysis) can provide some insights into why a single ODE can be sufficient to describe protein translation in the models of interest.

Extended S-System modelling of protein modification. After translation, proteins may undergo protein complex formation, protein stabilisation and/or protein degradation, amongst other processes (*i.e.* post-translational protein modification). Following [15, 44, 45], the formation of a protein complex C composed of N proteins $\{P_1, P_2, \dots, P_N\}$ can be modelled by the equation

$$\frac{dC}{dt} = aP_1P_2 \dots P_N - bC, \quad (12)$$

where a and b denote the rates of protein-protein association and complex degradation, respectively.

On the other hand, if we assume for example, that protein P is translated from gene G , stabilised by P_S and degraded by P_D , then following [15, 44, 45], these protein-mediated stabilisation and degradation processes can be modelled by the equation

$$\frac{dP}{dt} = aG - bP + c_S PP_S - c_D PP_D, \quad (13)$$

where parameters a , b , c_S and c_D represent the rates of translation, degradation, stabilisation and protein-mediated degradation, respectively.

Note that Eq (12) is already in the extended S-System form with the exponents $g_{i,j}$ and $h_{i,j,k}$ set to unity. Likewise, it can also be clearly seen that Eq (13) can be cast in the form of Eq (2) through appropriate choices of coefficients and exponents.

Extended S-System formulations of existing plant clock models

Using the framework outlined above, we constructed the extended S-System versions of four well-established plant clock models of varying complexity—JL2005 [11], JL2006 [43], AP2012 [44] and KF2014 [45], where we have used the initials of the leading author's first and last names followed by the year of publication to name the models. Each of these models employed the conventional Michaelis-Menten based modelling approach. Here, we append the notation 'S' to each plant clock model to denote its S-System variant (*e.g.* JL2005S denotes the extended S-System formulation of JL2005). In order to assess the degree to which these variants could reproduce the dynamics of the standard ODE representations, each extended S-System formulation was fitted to a synthetic dataset generated by the original model (*training data*) [12]. The out-of-sample error was then evaluated by scoring the extended S-System model against a second, distinct synthetic dataset also generated by the original model (*validation data*).

In addition, to further probe the predictive capacity of our formalism, we developed an extended S-System variant, MF2016KS, of a more recent clock model, MF2016K [15], which was fitted to an experimental training dataset. We also constructed a version of this model,

MF2016K Sorig, using the standard S-System formulation of Eq (1), and fitted it to the same dataset in order to assess the extent to which the extended formalism improved data-fitting. Mirroring the method used for synthetic data, the out-of-sample error for each model was evaluated with an experimental validation dataset.

The original ODE formulations for each model are given as eqs. (S1.1) (JL2005), eqs. (S1.3) (JL2006), eqs. (S1.5) (AP2012), eqs. (S1.7) (KF2014) and eqs. (S1.9) (MF2016K) of section 1 in S1 Text.

Fitting to synthetic data

For each model, we first generated timeseries for all circadian genes and proteins under transition from a 12L:12D light-dark cycle (*i.e.* alternating intervals of 12 hours of light and 12 hours of dark) to a constant light regime (LL). The parameters of the corresponding extended S-System formulation were then fitted to this synthetic training set by minimising the weighted mean squared error (WMSE) between the simulated and generated timeseries, *i.e.* by finding

$$\hat{\Theta}_{LL}^S = \underset{\Theta}{\operatorname{argmin}} W(\mathbf{X}_{LL}, \hat{\mathbf{X}}_{LL}(\Theta)), \quad (14)$$

where

$$W(\mathbf{X}_{LL}, \hat{\mathbf{X}}_{LL}(\Theta)) = \frac{1}{N_G} \sum_{i=1}^{N_G} W_i(X_i^{LL}, \hat{X}_i^{LL}(\Theta)) \quad (15)$$

and

$$W_i(X_i^{LL}, \hat{X}_i^{LL}(\Theta)) = \frac{1}{N_T} \sum_{j=1}^{N_T} \left(\frac{X_i^{LL}(t_j) - \hat{X}_i^{LL}(t_j, \Theta)}{A_i} \right)^2 \quad (16)$$

with

$$A_i = \max_{1 \leq j \leq N_T} X_i^{LL}(t_j). \quad (17)$$

The total WMSE, W , is the sum of the individual WMSEs, W_i , computed for each of the N_G circadian components in the plant clock model, for a given parameter set Θ . As different genes/proteins have different amplitudes, the weights A_i in the expression for W_i normalise each timeseries to its maximum value, in order to mitigate bias in the optimisation procedure when fitting the model parameters. In Eqs (15) and (16),

$$\mathbf{X}_{LL}(t) = (X_1^{LL}(t), \dots, X_{N_G}^{LL}(t)) \quad (18)$$

represent the timeseries generated from the original plant clock model in the simulated 12L:12D→LL transition, whilst

$$\hat{\mathbf{X}}_{LL}(t, \Theta) = (\hat{X}_1^{LL}(t, \Theta), \dots, \hat{X}_{N_G}^{LL}(t, \Theta)) \quad (19)$$

are the timeseries generated by the extended S-System variant in the same simulated transition for parameters Θ and N_T is the number of timeseries points used to score each circadian component. The minimisation was carried out using the MATLAB function `fminsearch`, which implements the Nelder-Mead simplex algorithm [46].

Next, for each model, we compared the dynamics of the original model and its extended S-System formulation under a different light condition—the transition from a 12L:12D light-dark cycle to constant dark (DD). To quantitatively assess the performance of the model on

this validation set, the total WMSE was calculated using Eqs (15) and (16) as

$$W(\mathbf{X}_{DD}, \hat{\mathbf{X}}_{DD}(\hat{\Theta}_{LL}^S)), \quad (20)$$

where

$$\mathbf{X}_{DD}(t) = (X_1^{DD}(t), \dots, X_{N_G}^{DD}(t)) \quad (21)$$

denotes the timeseries generated from the original model in the simulated 12L:12D→DD transition, and

$$\hat{\mathbf{X}}_{DD}(t, \Theta) = (\hat{X}_1^{DD}(t, \Theta), \dots, \hat{X}_{N_G}^{DD}(t, \Theta)) \quad (22)$$

is the corresponding timeseries generated by the extended S-System formulation for parameters Θ .

The parameter sets used to generate synthetic data for each model are listed in [S2 Table](#) (JL2005), [S5 Table](#) (JL2006), [S8 Table](#) (AP2012) and [S11 Table](#) (KF2014). We refer to these parameter sets as the nominal parameter values in each case. The MATLAB files used to generate synthetic data from each of these clock models can be downloaded at <https://github.com/mathiasfoo/essystemplantcircadian>.

Fitting to experimental data

In [15], two models of the plant circadian clock were developed—the full model, labelled MF2016, and the reduced kernel model, labelled MF2016K. The kernel model describes the core genetic circuitry that is responsible for generating wild-type behaviour of the plant circadian clock. Both the original (MF2016KSorig) and extended S-System (MF2016KS) formulations of this model were fitted to experimental data recorded in a 12L:12D→LL transition by finding the parameter set

$$\hat{\Theta}_{LL}^E = \underset{\Theta}{\operatorname{argmin}} W(\mathbf{D}_{LL}, \hat{\mathbf{X}}_{LL}(\Theta)), \quad (23)$$

calculated using Eqs (15) and (16), where

$$\mathbf{D}_{LL}(t) = (D_1^{LL}(t), \dots, D_{N_G}^{LL}(t)) \quad (24)$$

denotes the experimental timeseries and

$$\hat{\mathbf{X}}_{LL}(t, \Theta) = (\hat{X}_1^{LL}(t, \Theta), \dots, \hat{X}_{N_G}^{LL}(t, \Theta)) \quad (25)$$

represent the timeseries generated by MF2016KS for parameters Θ in a simulated 12L:12D→LL transition. Like in the case of synthetic data, minimisation was carried out using `fminsearch`. We note that in fitting MF2016KSorig to experimental data, we made a minor amendment to the light forcing term in [Eq \(3\)](#), setting the minimum value of $L_I(t)$ to 0.001 instead of 0. This was because in the original S-System formulation, production terms involving the expression $L_I(t)^{g_{ij}}$ become undefined if $L_I(t) = 0$ and $g_{ij} < 0$ (see [Eq \(1\)](#) above and eqs. (S1.11) in [S1 Text](#)).

Replicating the approach used for synthetic data-fitting, the validation goodness-of-fit was then calculated using experimental timeseries recorded in a 12L:12D→DD transition as

$$W(\mathbf{D}_{DD}, \hat{\mathbf{X}}_{DD}(\hat{\Theta}_{LL}^E)), \quad (26)$$

where

$$\mathbf{D}_{DD}(t) = (D_1^{DD}(t), \dots, D_{N_G}^{DD}(t)) \quad (27)$$

denotes the experimental timeseries and

$$\hat{\mathbf{X}}_{DD}(t, \Theta) = (\hat{X}_1^{DD}(t, \Theta), \dots, \hat{X}_{N_G}^{DD}(t, \Theta)) \quad (28)$$

is the corresponding simulation of MF2016KS or MF2016KSorig for parameters Θ .

The experimental data used for training and validation is presented as DataSet S1 and Table S1 of the Supporting Information in [15]. Note that not all experimental data was available in the literature for all circadian genes in all light conditions (for example, there was no data available for LHY protein under LL and DD conditions—see [15, 45]). To address this issue, the authors in [15] had used a data processing approach that combined synthetic and experimental data to produce timeseries for unavailable components. In this work, we use the same processed data from [15] to implement the parameter optimisation procedure described above. We also compare the fits to this data obtained with the two S-System models of the kernel structure, MF2016KS and MF2016KSorig, with the fit obtained previously in [15] using the Michaelis-Menten model formulation, MF2016K. The MATLAB files used for implementing MF2016K, MF2016KS and MF2016KSorig can be downloaded at <https://github.com/mathiasfoo/essystemplantcircadian>.

Assessing relative quality of fit using the AIC. In order to quantify the relative quality of the fits to the experimental training data obtained with MF2016KS, MF2016KSorig and MF2016K, we employed the widely-used Aikake Information Criterion (AIC), which calculates the best approximating model to a given dataset with respect to Kullback-Leibler information loss [47, 48]. For a given model, the AIC is defined as

$$\text{AIC} = -2 \ln(\hat{\mathcal{L}}) + 2K, \quad (29)$$

where $\ln(\hat{\mathcal{L}})$ is the maximised log-likelihood and K is the total number of estimated parameters. For the models considered here, since optimal parameter estimates were obtained by minimising a weighted least squares cost function, it can be shown that

$$\ln(\hat{\mathcal{L}}) = -\frac{N_G N_T}{2} \ln(2\pi + 1) - N_T \sum_{i=1}^{N_G} \ln(A_i) - \frac{N_G N_T}{2} \ln \left(W \left(\mathbf{D}_{LL}, \hat{\mathbf{X}}_{LL}(\hat{\Theta}_{LL}^E) \right) \right), \quad (30)$$

where the A_i s (defined in Eq (17)) are the cost function weights and $\hat{\Theta}_{LL}^E$ (defined in Eq (23)) is the parameter set that minimises the cost function [49].

Writing AIC_i for the AIC value of the i th model, we rank the three models of interest by calculating the AIC differences

$$\Delta_i(\text{AIC}) = \text{AIC}_i - \min_{1 \leq i \leq 3} \text{AIC}_i, \quad (31)$$

and the corresponding Aikake weights, defined below:

$$w_i(\text{AIC}) = \frac{\exp \left(-\frac{1}{2} \Delta_i(\text{AIC}) \right)}{\sum_{i=1}^3 \exp \left(-\frac{1}{2} \Delta_i(\text{AIC}) \right)}. \quad (32)$$

Weight $w_i(\text{AIC})$ can be interpreted as the probability that the i th model is the best (in the sense of minimising K-L information loss), given the data and set of candidate models. Furthermore, the strength of evidence in favour of model i over model j is quantified by the ratio $w_i(\text{AIC})/w_j(\text{AIC})$ [47–50].

Finally, since N_G , N_T and A_i in Eq (30) are fixed across the models, it follows that the AIC differences and Aikake weights can be computed from the following simplified expression for the AIC value of a given model

$$\text{AIC} = N_G N_T \ln(W(\mathbf{D}_{LL}, \hat{\mathbf{X}}_{LL}(\hat{\Theta}_{LL}^E))) + 2(K_\Theta + 1). \quad (33)$$

In the above, K_Θ denotes the number of model parameters that are optimised using the weighted least squares cost function to calculate $\hat{\Theta}_{LL}^E$.

Variability in optimised parameter values

In order to quantify the parameter variability associated with our optimisation procedure when fitting extended S-System models to synthetic/experimental data, we performed the following analysis. For each model, following an initial parameter optimisation run, five further runs were executed in which the initial parameters were sampled from a multivariate normal distribution with its mean set to the optimal values of the initial run and its covariance matrix set to the identity matrix. Parameter sampling was carried out using the MATLAB function `mvnrnd`. During the search process, if no further improvement was observed in the cost function (*i.e.* the WMSE value of the fit), this was taken to indicate the presence of a local minimum and the optimisation run was terminated. For each model, six parameter sets that could reproduce the synthetic/experimental data were thus generated in this manner. To obtain a robust measure of the variability of a given parameter θ , we then computed its normalised Median Absolute Deviation (nMAD), using the following equation:

$$\text{nMAD}(\theta) = \frac{\text{median}(|\theta_i - \tilde{\theta}|)}{|\tilde{\theta}|}. \quad (34)$$

In the above, θ_i is the optimal value of θ obtained for the i th run and $\tilde{\theta} = \text{median}(\theta_i)$.

Results

Extended S-System representations of the models

For each of the five plant clock models considered in this study, our extended S-System formulations of the corresponding GRN circuits are presented as eqs. (S1.2) (JL2005S), eqs. (S1.4) (JL2006S), eqs. (S1.6) (AP2012S), eqs. (S1.8) (KF2014S) and eqs. (S1.10) (MF2016KS) of section 1 in [S1 Text](#).

Fits to synthetic data

The parameter values yielding the best fits of each model to the corresponding training dataset are given in [S3 Table](#) (JL2005S), [S6 Table](#) (JL2006S), [S9 Table](#) (AP2012S) and [S12 Table](#) (KF2014S). Below, we discuss the performance of each optimised model in turn.

JL2005. The simplest plant clock model we consider here is JL2005 ([Fig 1A](#)), which only comprises two circadian genes, *LHY/CCA1* and *TOC1*. In [\[11\]](#), this circuit is described with seven ODEs, in which one equation is used exclusively to represent the interaction of light with a light-sensitive protein, and is therefore decoupled from the others. Since this light-sensitive protein is not part of the core plant clock, we omit it from the timeseries and heatmap plots used to present the fitting results below.

The fits of JL2005S to the training and validation sets are shown in [Fig 1B and 1C](#). In [Fig 1B](#), we plot the timeseries, while in [Fig 1C](#), we plot the corresponding heatmaps. For the heatmaps, each gene or protein is represented by two rows, where the top and bottom rows

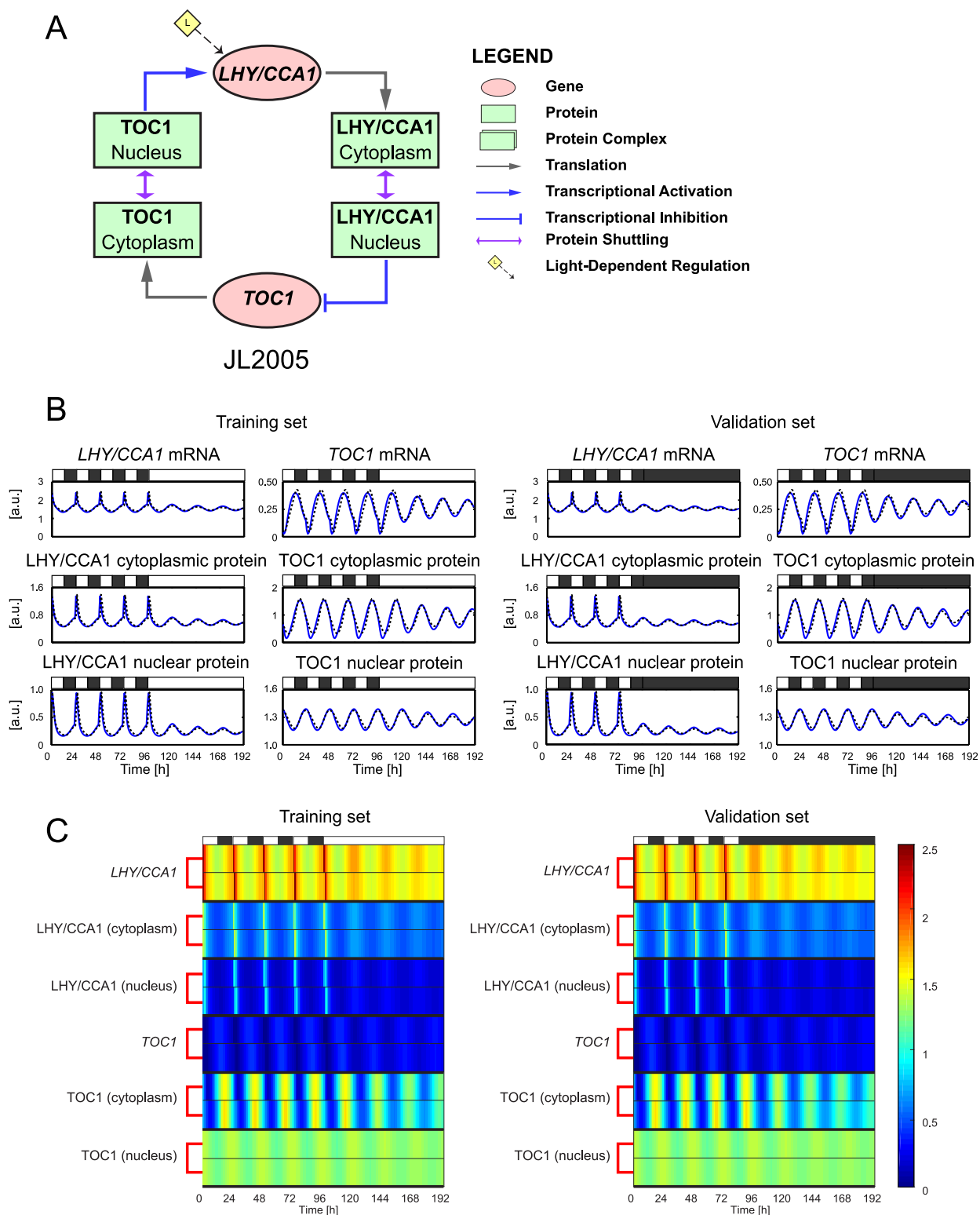


Fig 1. The JL2005S plant clock model—Fits to synthetic data. **A:** Regulatory circuit diagram for JL2005 [11]. Genes and proteins are represented as ovals and rectangles respectively. Grey solid lines represent translation. Blue solid lines with arrow heads (resp. bar heads) represent transcriptional activation (resp. inhibition). Double headed arrows denote protein shuttling between the cytoplasm and nucleus. The yellow diamond denotes light-dependent regulation. **B:** Comparison between expression timeseries in JL2005 (blue solid lines) and its extended S-System formulation JL2005S (black dashed lines). JL2005S was optimised to training data generated from JL2005 in a simulated 12L:12D → LL transition

(left panels). For validation, the behaviour of the models was then evaluated in a simulated 12L:12D \rightarrow DD transition (right panels). White and black bars represent intervals of light and dark. C: Heatmap representation of the timeseries shown in B. Each gene/protein is grouped together, with the JL2005 timeseries plotted in the top row of each group and the JL2005S timeseries plotted in the bottom row.

<https://doi.org/10.1371/journal.pcbi.1007671.g001>

represent the expression dynamics generated by JL2005 and JL2005S respectively. The timeseries and heatmaps demonstrate excellent agreement between the two models, as quantified by the small WMSE values given in [S4 Table](#). Both models attain peaks and troughs at the same times with near-identical amplitudes, demonstrating the capacity of the extended S-System formulation to reproduce the dynamics of the original equations with a simpler model structure. We also note that the model parameters are quite tightly constrained (*i.e.* have low variability across optimisation runs), with all parameters having nMAD values less than 0.25 (see [S12A Fig](#) and [S3 Table](#)).

JL2006. The second plant clock model, JL2006 was constructed by expanding JL2005 from a two-gene to a five-gene circuit [43] (see [Fig 2A](#)). Note that in JL2006, there are two *speculative* genes labelled *X* and *Y*, reflecting the fact that when modifying JL2005 to better fit experimental data, Locke *et al.* proposed that there should be a gene (*X*) that acts as an intermediate genetic component between *TOC1* and *LHY/CCA1*, and another gene (*Y*) that acts as an intermediate genetic component between *LHY/CCA1* and *TOC1*. The identity of these two genes was eventually discovered and they will be discussed when we consider AP2012 below. JL2006 comprises 16 ODEs. As with JL2005, we omit the light-sensitive protein from the presentation of the fitting results, as it is not part of the core circadian clock.

[Fig 2B and 2C](#) show the fits of JL2006S to synthetic data in the form of heatmaps (the corresponding WMSE values are given in [S7 Table](#)). Because the expression amplitude of each gene and protein differs significantly, we normalised each timeseries by its maximum value when plotting the heatmaps, yielding a maximum relative amplitude of one (the unnormalised timeseries are shown in [S1](#) and [S2 Figs](#)). For the training dataset, excellent agreement was observed between the two models, with both formulations attaining peaks and troughs at very similar times with similar amplitudes. Furthermore, JL2006S reproduces the acute light responses of the *Y* gene and *Y* protein that occur at dawn in light-dark cycles. For the validation dataset, both models also show good agreement, with the exception of *Y* mRNA and *Y* protein in DD conditions, where higher amplitudes are observed for JL2006S.

In terms of parameter variability, it can be seen in [S12B Fig](#) and [S6 Table](#) that similarly to JL2005S, the parameters of JL2006S are fairly well constrained, with the majority (56/61 parameters) having nMAD values less than 0.5, and all parameters having nMAD values less than 0.8.

AP2012. Although the extended S-System approach was very successful in reproducing the circadian dynamics of JL2005 and JL2006, these two plant clock models are characterised by simple transcription and translation mechanisms, where post-translational processes such as protein complex formation, protein stabilisation and protein degradation are not considered. The capacity of the extended S-System approach to describe these more complicated protein modification steps was assessed by applying it to the third plant clock model, AP2012 [44].

In AP2012, the previously unknown gene *Y* in JL2006 had been identified as *GI*. The other unknown component of JL2006, gene *X*, had been removed on the basis of new experimental work indicating that *TOC1* protein was a transcriptional *inhibitor* of *LHY/CCA1* [44, 51, 52], rather than a transcriptional *activator* as had been initially assumed (this change in *TOC1* function was also predicted by Boolean modelling in [12]). Modifying the regulation of *LHY/CCA1* by *TOC1* in accordance with these findings resulted in the new model being able to fit a

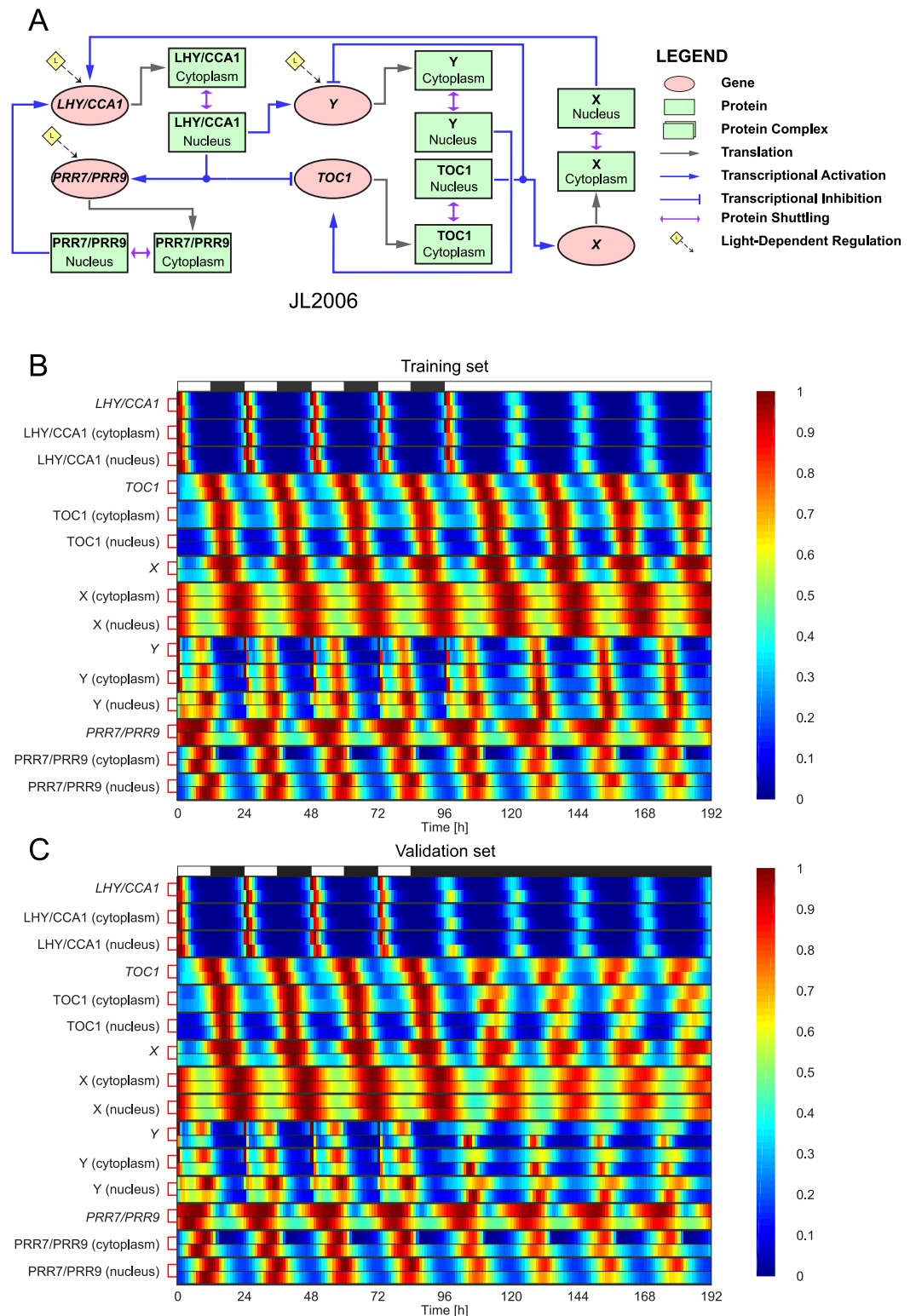


Fig 2. The JL2006S plant clock model—Fits to synthetic data. A: Regulatory circuit diagram for JL2006 [43]. The same symbols were used as in Fig 1. B–C: Comparison between expression timeseries in JL2006 and its extended S-System formulation JL2006S. JL2006S was optimised to training data generated from JL2006 in a simulated 12L:12D → LL transition (B). For validation, the behaviour of the models was then evaluated in a simulated 12L:12D → DD transition (C). Timeseries are presented as heatmaps in which each gene/protein is grouped together, with the top and bottom rows in each group showing JL2006 and JL2006S expression levels, respectively. To aid visualisation, each timeseries has been normalised to its maximum value. White and black bars represent intervals of light and dark.

<https://doi.org/10.1371/journal.pcbi.1007671.g002>

broader range of experimental data, leading to the removal of gene *X* from the model (and its subsequent extensions).

The circuit diagram for AP2012 is shown in Fig 3A. AP2012 is described by 28 ODEs. Heatmaps comparing the normalised dynamics of the model and its extended S-System formulation are shown in Fig 3B and 3C, with the corresponding unnormalised timeseries given in S3 and S4 Figs. In addition to the light-sensitive protein, we also omit the heatmap and timeseries of all COP1-related proteins in our plots, as these are not part of the core plant circadian network.

Like JL2005S and JL2006S, AP2012S displays excellent agreement on the training dataset. For the validation dataset, despite the simulated genes/proteins in AP2012S having similar peak and trough phases to their counterparts in AP2012, and a correspondingly small fitting error (cf. S10 Table), the extended S-System formulation generates an oscillation in DD that decays faster than the original formulation.

The parameter variability analysis of AP2012S shows that the optimisation process yields fairly constrained fits for this model also: the vast majority of parameters (96/115) have nMAD values less than 0.5 and the largest nMAD value is ≈ 0.94 (see S12C Fig and S9 Table).

KF2014. The fourth model, KF2014 [45], is the most comprehensive plant clock model available to date: its circuit diagram is shown in Fig 4A. KF2014 was constructed and validated against a large number of experiments reported in the literature (approximately 800 timeseries datasets spanning 59 published papers [45]). This results in the model being able to reproduce experimental findings across a broad range of different conditions. KF2014 is described by 35 ODEs. Heatmaps comparing the normalised dynamics of KF2014 and KF2014S are shown in Fig 4B and 4C (for the corresponding unnormalised timeseries, see S5 and S6 Figs). Like AP2012, the light-sensitive protein and all COP1-related proteins are omitted in our timeseries and heatmap plots.

For the training data, the heatmaps and timeseries indicate good agreement between the expression dynamics of the two models. For the validation set, both models show good agreement for the majority of components, with the exception of ELF3/ELF4 complex, *LUX* mRNA, *LUX* protein, *GI* mRNA, *GI* cytoplasmic protein and *GI* nuclear protein (cf. S6 Fig and S13 Table). However, the difference lies mainly in the amplitudes—the timing of peak and trough expression is similar in both models.

Finally, the variability in optimised parameter values is comparable to that observed for AP2012S, with a similar proportion of parameters (127/152) having nMAD values less than 0.5 and the maximum nMAD value being ≈ 0.94 (see S12D Fig and S12 Table).

Fits to experimental data

MF2016K. Here, we assess the ability of the extended S-System formulation to reproduce experimental data, where the modelled interaction between circadian genes follows the kernel version of the ODE system developed in [15]. The kernel model MF2016K shown in Fig 5A is described by 24 ODEs. Fig 5B and 5C show heatmaps comparing the normalised expression timeseries of the experimental datasets to the fits obtained previously with MF2016K in [15] and the fits obtained in this study with the extended/original S-System formulations of the model (MF2016KS/MF2016KSorig). The unnormalised timeseries for all three models are plotted in S7 and S8 Figs. The parameter sets yielding the best fits to experimental data are presented in S14 Table (MF2016KS), S16 Table (MF2016KSorig) and S17 Table (MF2016K), whilst the corresponding WMSE values are given in Table 1 (all models) and S15 Table (detailed cost breakdown for MF2016KS).

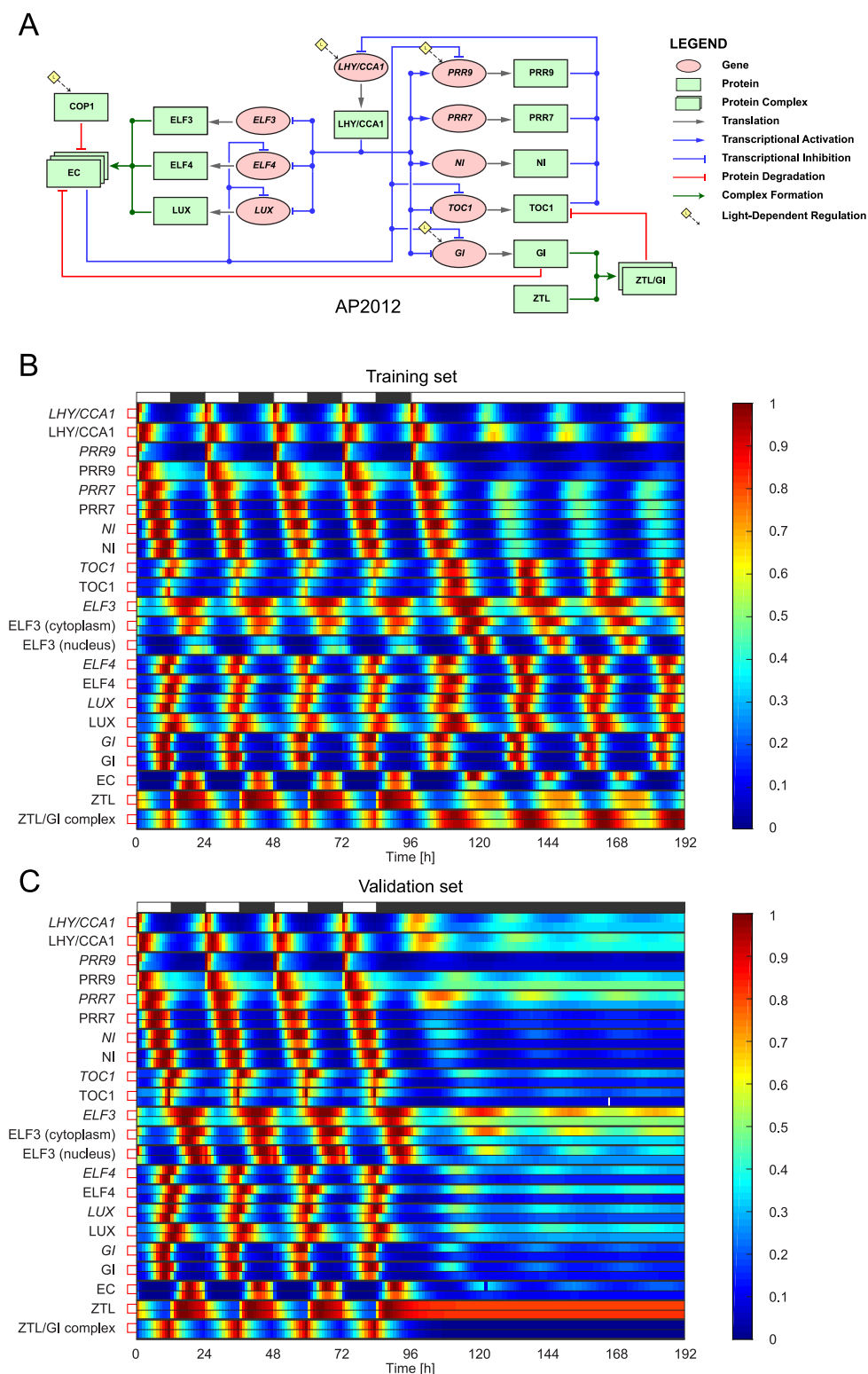


Fig 3. The AP2012S plant clock model—Fits to synthetic data. A: Regulatory circuit diagram for AP2012 [44]. The same symbols were used as in Fig 1, with the addition of green solid lines representing complex formation. B-C: Comparison between expression timeseries in AP2012 and its extended S-System formulation AP2012S. AP2012S was optimised to training data generated from AP2012 in a simulated 12L:12D \rightarrow LL transition (B). For validation, the behaviour of the models was then evaluated in a simulated 12L:12D \rightarrow DD transition (C). Timeseries are presented as heatmaps in which each gene/protein is grouped together, with the top and bottom rows in each group showing AP2012 and AP2012S expression levels, respectively. To aid visualisation, each timeseries has been normalised to its maximum value. White and black bars represent intervals of light and dark.

<https://doi.org/10.1371/journal.pcbi.1007671.g003>

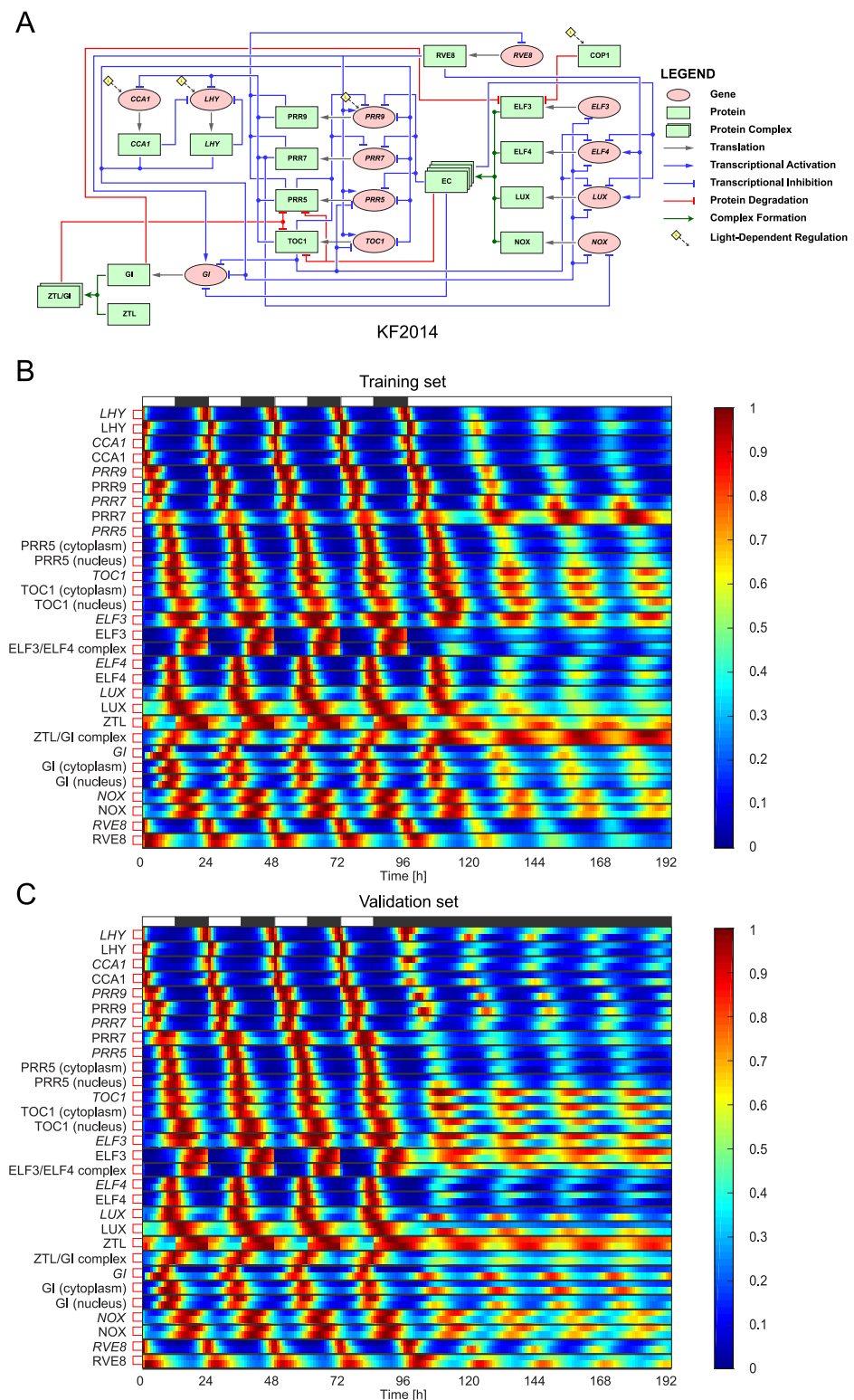


Fig 4. The KF2014S plant clock model—Fits to synthetic data. A: Regulatory circuit diagram for KF2014 [45]. The same symbols were used as in Fig 1. B-C: Comparison between expression timeseries in KF2014 and its extended S-System formulation KF2014S. KF2014S was optimised to training data generated from KF2014 in a simulated 12L:12D → LL transition (B). For validation, the behaviour of the models was then evaluated in a simulated 12L:12D → DD transition (C). Timeseries are presented as heatmaps in which each gene/protein is grouped together, with the top and bottom rows in each group denoting KF2014 and KF2014S expression levels, respectively. To aid visualisation, each timeseries has been normalised to its maximum value. White and black bars represent intervals of light and dark.

<https://doi.org/10.1371/journal.pcbi.1007671.g004>

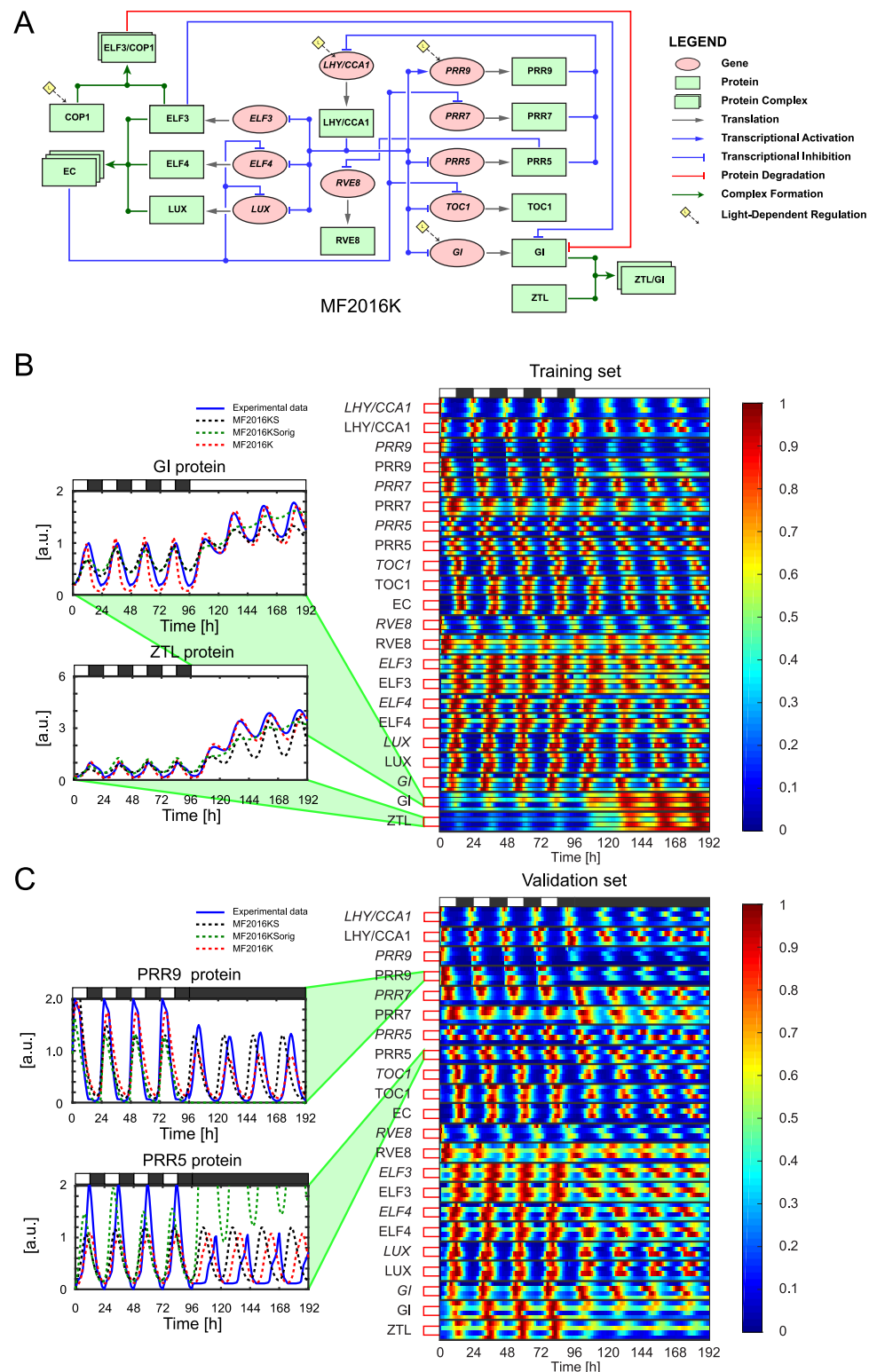


Fig 5. The MF2016KS, MF2016K and MF2016KSorig plant clock models—Fits to experimental data. A: Regulatory circuit diagram for MF2016K [15]. The same symbols were used as in Fig 1. B-C: Comparison between experimental expression timeseries and the corresponding simulations generated by the extended S-System formulation MF2016KS, the Michaelis-Menten formulation MF2016K and the original S-System formulation MF2016KSorig. The models were optimised to experimental data recorded in a 12L:12D → LL transition (B). For validation, the behaviour of each

model in a simulated 12L:12D \rightarrow DD transition was then evaluated against experimental data recorded in the same conditions (C). Timeseries are presented as heatmaps in which each gene/protein is grouped together, with the first row in each group showing experimental expression levels and the second, third and fourth rows showing simulated expression levels from models MF2016KS, MF2016K and MF2016KSorig, respectively. To aid visualisation, each timeseries has been normalised to its maximum value. White and black bars represent intervals of light and dark. In order to highlight the differences in predictive capacity between the three models, (B) shows the unnormalised expression timeseries for GI and ZTL proteins, while (C) shows the unnormalised expression timeseries for PRR9 and PRR5 proteins.

<https://doi.org/10.1371/journal.pcbi.1007671.g005>

For the training dataset, the heatmaps and timeseries indicate good agreement between both MF2016KS and the experimental data, with the extended S-System model yielding very similar expression dynamics to the Michaelis-Menten formulation (see Fig 5B and S7 Fig). The comparable performance between the models in reproducing the experimental data is reflected by their near-identical WMSE values, although MF2016KS has fewer parameters than MF2016K (see Table 1). The standard S-System formulation, MF2016KSorig, did not reproduce the behaviour of either GI or ZTL protein (see Fig 5B and S7 Fig). Thus, despite being able to adequately reproduce the dynamics of other genes and proteins, MF2016KSorig has a larger WMSE value than the other two models.

The AIC values in Table 1 quantify this comparison in model performance, with MF2016KS, MF2016K and MF2016KSorig yielding Aikake weights of $w_{KS} = 0.9981$, $w_K = 0.0019$ and $w_{KSorig} = 0$, respectively. These weights exclude MF2016KSorig as a viable candidate for the best model (in the sense of K-L divergence) and imply strongly that the best model is MF2016KS (the evidence ratio $\frac{w_{KS}}{w_K}$ indicates that it is 530 times more likely to be so than MF2016K).

For the validation dataset, although the simulated MF2016KS waveforms (and hence amplitudes and periods) of most circadian components are close to those of the experimental timeseries, peak expression occurs earlier in the model following release from LD into DD (see Fig 5C and S8 Fig). In addition, for ZTL protein, the MF2016KS timeseries has a larger amplitude than the experimental expression profile. MF2016K yields a better fit to the data in this case (particularly to ZTL protein), as reflected by its lower WMSE value (*i.e.* its lower out-of-sample error). MF2016KSorig has a high out-of-sample error, predominately due to some simulated PRR9, PRR7 and PRR5 components having much higher amplitudes than their experimental counterparts in DD (see Fig 5C and S8 Fig).

The parameter variability analysis of MF2016KS mirrors the general trend observed in the fits of the larger models to synthetic data. The overwhelming majority of parameters

Table 1. Ranking model fits to experimental data. Here, N_G and N_T are the number of gene/protein timeseries and the number of timepoints per timeseries used for fitting, respectively. $W(\mathbf{D}_{LL}, \hat{\mathbf{X}}_{LL}(\hat{\Theta}_{LL}^E))$ is the weighted mean squared error (WMSE) of the best fit to the training data and $W(\mathbf{D}_{DD}, \hat{\mathbf{X}}_{DD}(\hat{\Theta}_{LL}^E))$ is the corresponding WMSE value of the fit to the validation data. $\Delta_i(\text{AIC})$ and $w_i(\text{AIC})$ denote the AIC differences and Aikake weights for each model, respectively, calculated using Eqs (31)–(33).

	Model		
	MF2016KS	MF2016KSorig	MF2016K
N_G	24		
N_T	194		
K_Θ	72	78	77
$W(\mathbf{D}_{LL}, \hat{\mathbf{X}}_{LL}(\hat{\Theta}_{LL}^E))$	0.03647	0.04461	0.03649
$W(\mathbf{D}_{DD}, \hat{\mathbf{X}}_{DD}(\hat{\Theta}_{LL}^E))$	0.13557	1.70248	0.05629
$\Delta_i(\text{AIC})$	0	950.03	12.55
$w_i(\text{AIC})$	0.99812	0	0.00188

<https://doi.org/10.1371/journal.pcbi.1007671.t001>

(63/72) have nMAD values less than 0.5 and the largest nMAD value is ≈ 0.93 (see [S12E Fig](#) and [S14 Table](#)).

The aggregated protein model—Linear systems analysis

As noted earlier, the later plant clock models (e.g. AP2012 and KF2014) used a single equation to represent the production of TF, except in cases where experimental data was available that distinguished between cytoplasmic and nuclear forms (e.g. ELF3 and GI in [\[53\]](#) and PRR5, TOC1 and GI in [\[54\]](#)). This was in contrast to the paired equations used to represent protein shuttling between cytoplasm and nucleus that were employed in the earlier models (e.g. JL2005 and JL2006).

To understand why the practice of aggregating protein forms did not appear to adversely affect the predictive capacity of the later models, we used sine-sweeping—a frequency response analysis method from linear systems theory [\[55\]](#)—to approximate the transfer functions describing the production of TF in the models that explicitly describe transport between cytoplasmic and nuclear compartments. Sine-sweeping is widely used in the field of system identification to obtain simplified empirical transfer functions. The method is applicable when the system of interest displays linear dominant behaviour (*i.e.* when the response to a sinusoidal input signal is itself sinusoidal, with the same input frequency and some phase shift). For a linear dominant system, sine-sweeping provides a means of obtaining an approximate, simplified linear model in the event that the linearisation of the full nonlinear model is not straightforward. When we applied the method to the plant clock models incorporating protein shuttling, JL2005 and JL2006, we observed linear dominant behaviour in each case, thereby validating the approach. Here, we present the results obtained for JL2006 (similar results were obtained for JL2005).

In applying sine-sweeping to the protein shuttling mechanism, the system input is the mRNA expression timeseries and the system output is the resulting nuclear protein expression timeseries (see [Fig 6A](#)). We then drive the system with sinusoidal input signals with frequencies in the range 0.01 rad/h to 2 rad/h. From systems theory, it follows that if the system is linear (or linear dominant), the output obtained for a sinusoid of given frequency will also be sinusoidal with the same frequency, but with a scaled amplitude and phase shift. By calculating the amplitude scaling and phase shifts over the frequencies of interest, we obtain a Bode plot. This in turn allows us to approximate the transfer function of the system, which enables the system's response to any input signal (*i.e.* any mRNA expression timeseries) to be estimated (see section 2 in [S1 Text](#) and [S9 Fig](#)). The order N of the transfer function is of particular importance in our case, as it specifies the minimum number of linear differential equations required to represent the system. N can be approximated from the Bode plot by exploiting the fact that for an N th order transfer function, the slope of the magnitude plot at the corner frequency is $(-20 \times N)$ dB/decade and the phase shift at this frequency is $(-45 \times N)^\circ$ (the corner frequency is defined as the frequency at which the magnitude plot has decreased by 3dB from its plateau level—for more details, see [\[56\]](#)).

The protein shuttling equations in JL2006 have the general form given in [Eq \(8\)](#) but with nonlinear rather than linear degradation terms (*cf.* eqs. (S1.3) in [S1 Text](#)). The Laplace transforms of the linearised protein shuttling equations in JL2006 yield second-order transfer functions, as detailed in section 3 of [S1 Text](#). For each gene, we would therefore expect the Bode plot obtained with sine-sweeping to have a slope of -40 dB/decade at the corner frequency and a corresponding phase shift of -90° . However, as can be seen in [Fig 6B–6F](#), for all the circadian genes, the phase shift at the corner frequency is instead close to -45° and the corresponding slope in the magnitude plot is close to -20 dB/decade (with the exception of

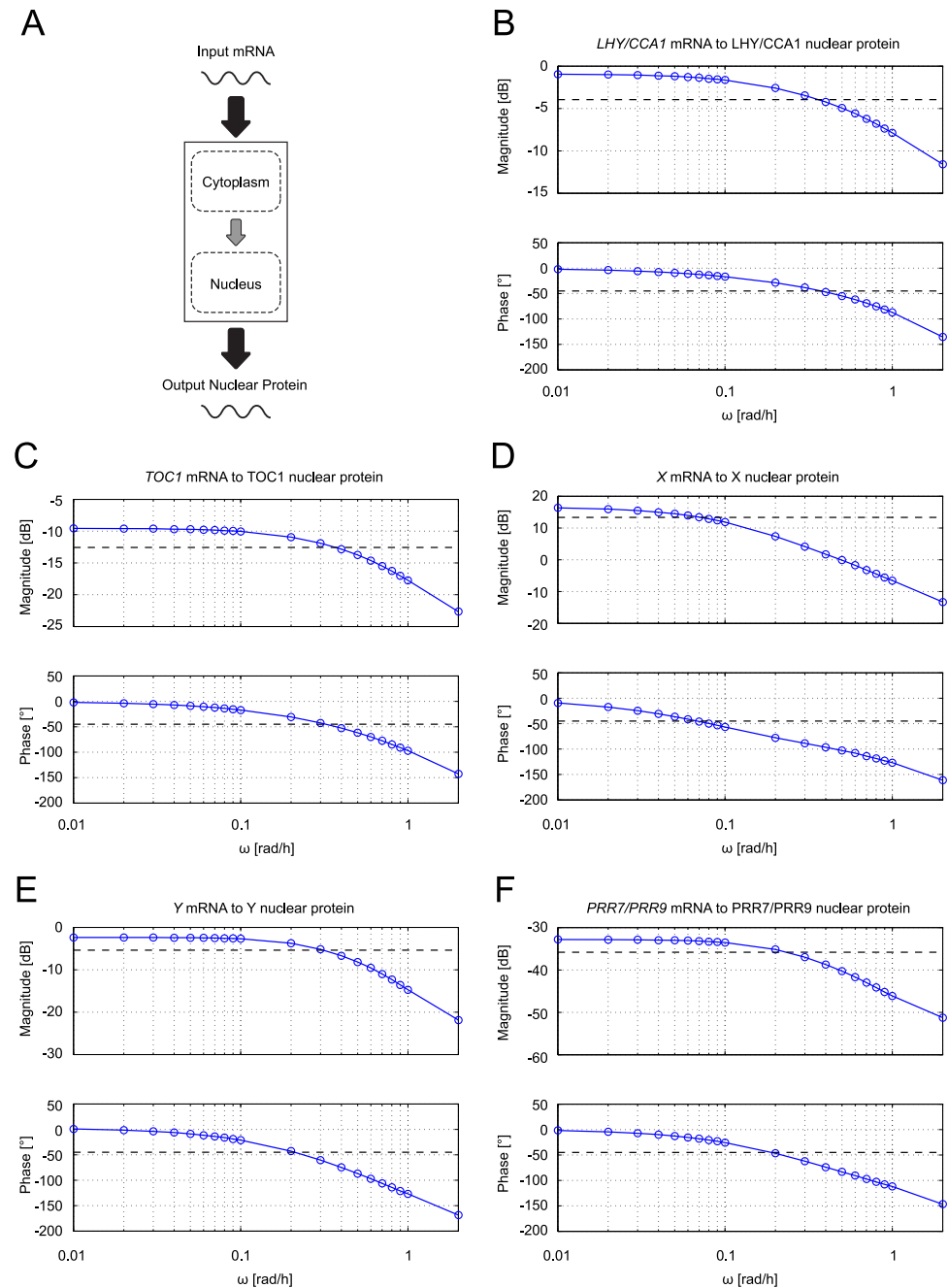


Fig 6. Frequency response analysis of protein shuttling in JL2006. **A:** For each circadian gene, a sinusoidal mRNA signal with varying frequencies is provided as the input to the system and the nuclear protein is observed as the resulting output. **B-F:** Bode plots obtained from sine-sweeping the five different circadian genes of the model. In each magnitude plot, the black dashed line represents the magnitude that is 3dB lower than that at $\omega = 0$. For each phase plot, the black dashed line represents a phase shift of -45° . For each gene, if the protein shuttling mechanism can be approximated by a first-order transfer function, then the intersections of the blue solid lines and the black dashed lines in both the magnitude and phase plots should occur at the same frequency. **B:** *LHY/CCA1* mRNA \mapsto *LHY/CCA1* nuclear protein. **C:** *TOC1* mRNA \mapsto *TOC1* nuclear protein. **D:** *X* mRNA \mapsto *X* nuclear protein. **E:** *Y* mRNA \mapsto *Y* nuclear protein. **F:** *PRR7/PRR9* mRNA \mapsto *PRR7/PRR9* nuclear protein.

<https://doi.org/10.1371/journal.pcbi.1007671.g006>

LHY/CCA1, for which the slope is approximately -10 dB/decade). Thus, the protein shuttling mechanism in each case appears to be quite well approximated with a first-order transfer function.

To understand the reason for this observed difference, we further analysed the protein equations used for JL2006. A second order system can be approximated by a first-order system if the two poles (*i.e.* the roots of the denominator of the transfer function) are far apart—in other words, if the system has one fast pole and one slow pole. Using gene *Y* as a case study, we found that the two poles are indeed far apart (see eq. (S3.5) in [S1 Text](#) and [S10 Fig](#)), with one pole having a value approximately 30 times greater than the other. As shown in [S11 Fig](#), this scenario results in the second-order transfer function quite closely resembling a first-order transfer function, which implies that a single, aggregated protein equation is sufficient to represent the protein translation pathway (see eqs. (S3.6) and (S3.7) in [S1 Text](#)).

Discussion

Extended S-System formulations can reproduce the dynamics of existing plant clock models and experimental data

In this study, we have investigated the use of a simplified modelling framework based on S-Systems to describe the behaviour of the plant circadian clock. We tested the efficacy of this new approach by constructing the extended S-System formulations of five different existing plant clock models. Four of these models—the extended S-System versions of JL2005 [[11](#)], JL2006 [[43](#)], AP2012 [[44](#)] and KF2014 [[45](#)—were optimised to synthetic training data generated from the original models, whilst the extended S-System version of the fifth model—MF2016K [[15](#)—was optimised to experimental training data. To assess predictive capacity, the goodness-of-fit obtained for each model on a validation dataset (*i.e.* one distinct from the training data) was then computed.

For the two simplest models considered, JL2005 and JL2006, very close agreement was observed between the extended S-System formulations and the original models for both the training and validation datasets (Figs [1](#) and [2](#), [S1](#) and [S2](#) Figs), with near-identical simulated and target expression timeseries in some of the model components. For AP2012 and KF2014, which incorporate a greater number of genetic components and more complex regulation mechanisms, excellent fits were again observed for the training data (Figs [3B](#) and [4B](#), [S3](#) and [S5](#) Figs). On validation data, whilst AP2012S gives a close match to AP2012 during the light-dark cycle, the extended S-System formulation generates a much more pronounced damping (with phase shifts in some components) following release into constant dark ([Fig 3C](#) and [S4 Fig](#)). Similarly to AP2012S, KF2014S generates timeseries that closely match KF2014 in the LD portion of the validation dataset, with a more pronounced deviation between the models observed following DD release ([Fig 4C](#) and [S6 Fig](#)).

Interestingly, despite AP2012S's poorer fit to validation data, a comparable predictive performance to AP2012 is observed when qualitatively modelling short- and long-period mutant phenotypes. This can be seen in [Table 2](#), which compares the predicted period phenotypes generated by the two model formulations against the corresponding experimentally-observed phenotype for a range of knockout/knockdown and overexpression mutants in different light conditions. Indeed, the table shows that both AP2012 and its extended S-System formulation correctly simulate 10/16 phenotypes (albeit not all the same ones), despite the latter not having been fitted to any mutant data.

In terms of experimental data-fitting, the results obtained for MF2016KS mirrored those of the more complex models on synthetic data. The extended S-System model generates expression timeseries that give good matches to the training data ([Fig 5B](#) and [S7 Fig](#)). Moreover, the

Table 2. Comparisons between simulated and experimental free-running period shifts $\delta\tau$ in different constant light conditions and genetic backgrounds for plant models AP2012 and AP2012S. For all light-mutant combinations considered, $\delta\tau$ was calculated as the difference between the mutant and wild-type periods: $\delta\tau = \delta\tau_{mut} - \delta\tau_{wt}$. Hence, $\delta\tau < 0$ corresponds to a short-period mutant (denoted by a – sign) and $\delta\tau > 0$ corresponds to a long-period mutant (denoted by a + sign). The study providing the experimental values used to calculate $\delta\tau$ is reported in the rightmost column in each case. In the above, Δ denotes knockout/knockdown, ‘OX’ denotes overexpression, ‘arr.’ denotes an arrhythmic oscillation, LL denotes constant light and DD denotes constant dark. Knockout/knockdown mutant behaviour was simulated by reducing the transcription rate of the target TF by 80%, with the exception of the ZTL mutant for which the protein production rate constant was decreased by 80% instead (there is no term for ZTL mRNA production in AP2012—see eqs. (S1.5) in S1 Text). Overexpression mutant behaviour was simulated by increasing the translation rate of the target TF two-fold. Simulated periods were calculated by using the MATLAB function `findpeaks` to obtain all the differences between two successive maxima in the circadian rhythm, and then averaging across all circadian gene components.

Mutant	Light condition	sgn($\delta\tau$): exp.	sgn($\delta\tau$): simulated		Citation
			(AP2012S)	(AP2012)	
$\Delta lhy/cca1$	LL	-	+	-	[57]
$\Delta toc1$	LL	-	-	-	[58]
$\Delta prr7$	LL	+	-	+	[59]
$\Delta prr9$	LL	+	+	-	[59]
$\Delta prr9prr7$	LL	+	+	+	[59]
Δztl	LL	+	+	+	[60]
Δgi	LL	-	-	-	[61]
$\Delta elf4$	LL	arr.	arr.	arr.	[62]
$\Delta elf3$	LL	arr.	arr.	arr.	[63]
Δlux	LL	arr.	arr.	arr.	[64]
ELF4-OX	LL	+	-	-	[65]
ELF3-OX	LL	+	-	-	[65]
Δgi	DD	arr.	-	-	[66]
$\Delta elf4$	DD	arr.	arr.	+	[67]
Δztl	DD	+	+	+	[68]
ELF3-OX	DD	+	-	-	[65]

<https://doi.org/10.1371/journal.pcbi.1007671.t002>

extended S-System formalism yielded a superior model in this case, compared to both the original Michaelis-Menten formulation and a model based on the standard S-System framework, as quantified by an AIC analysis (Table 1). Similarly to AP2012S and KF2014S, however, a greater discrepancy between MF2016KS and data is observed during the LD to DD transition, predominately characterised by a phase shift (Fig 5C and S8 Fig). The Michaelis-Menten model gives a better fit in this case, albeit with a larger number of parameters.

The extended S-System framework facilitates model development, network inference and synthetic circuit design

The results of our fits to synthetic and experimental data demonstrate that the extended S-System formulation is capable of yielding models with comparable predictive capacity to a set of canonical plant clock models. Furthermore, this approach confers several advantages for GRN modelling compared to the conventional Michaelis-Menten based framework that is predominately used in computational circadian biology.

The first such advantage relates to model development and expansion. The extended S-System formulation enables new interactions to be added easily, as depending on whether such an interaction affects the production, degradation or light regulation of the target component, it can simply be incorporated into the corresponding term of Eq (2) through an appropriate choice of coefficients, exponents and upper product/summation bounds. Conversely, interactions can also be removed in a straightforward manner.

The second advantage relates to network inference, which is the concomitant of the first advantage. In the extended S-System formalism, the type of regulation implemented by a given

network component is simply determined by the sign of the corresponding exponent $g_{i,j}$ in the first term of Eq (2), with $g_{i,j} < 0$ indicating an inhibitor and $g_{i,j} > 0$ indicating an activator. The regulation type can therefore be inferred together with all the other parameters specifying Eq (2) during data-fitting, without having to modify the production term. Indeed, for all the data-fitting presented here, no constraints were imposed on the $g_{i,j}$ s for the initial optimisations to training data. For synthetic data-fitting, the inferred patterns of activation and inhibition were checked against the corresponding model, whilst for optimisation to experimental data, the activation/inhibition pattern was checked against the experimental literature. In each case, almost all signs were correctly inferred. In the event that one or more signs were incorrectly identified, these were reversed and a further optimisation run was performed.

To put these two advantages into perspective, in order to add a new interaction to one of the plant clock models using the Michaelis-Menten model structure, we would need to specify *a priori* whether this was a positive or negative interaction, given the different functions used to represent activation and inhibition (cf. Eqs (5) and (6)) and then estimate the two new parameters associated with this interaction. If the interaction type was unknown, two separate optimisations would be required. On the other hand, using the extended S-System formulation, we would only require an estimate of the exponent associated with the new interaction (cf. Eq (2)) and if the interaction type were unknown, a single optimisation would be sufficient to infer it. In this vein, recent experimental work appears to imply a switch in $\text{LHY} \mapsto \text{PRR9}$ regulation from activation to inhibition [69]. Revising this interaction could be done in a model based on the extended S-System formalism in a very straightforward manner, by reversing the sign of the exponent associated with the interaction and then reoptimising parameters. In addition, the interaction type (activation or inhibition) that was predicted to give the best fit to data could be established by simply leaving the sign of the exponent unconstrained during the parameter optimisation process.

A third key advantage relates to the design of synthetic feedback control circuits for mitigating perturbations to GRNs (e.g. changes to steady-state expression levels associated with infection). Indeed, our previous work [70] has shown that using Michaelis-Menten based models to obtain accurate simulations of closed-loop control strategies requires *consistent* estimates of the Michaelis constants for all components affected by the control signal(s) (here, by consistent we mean multiple optimisation runs started from different initial conditions locate similar parameter values that reproduce the data). This is because variability in these estimates can lead to a simulated saturated response in the relevant components (i.e. if $P \gg K$ in Eqs (5) and (6)). This in turn can result in an inaccurate prediction of controller behaviour, as shown in Figure 3 of [70]. However, the propensity of Hill functions to generate a saturated response means that, in practice, optimisations to data can lead to large variations in the inferred values of the corresponding Michaelis constants—in other words, models utilising such functions can suffer from poor parameter identifiability.

Finally, although our previous analysis concurred with [71] that accurately estimating exponents in power-law based models can be challenging (see Figures 4 and 7 of [70]), it also showed that these estimates are more consistent than those obtained for the corresponding parameters in Michaelis-Menten based models. This suggests that the S-System formalism can mitigate the parameter identifiability issue, identifying the approach as a viable complementary modelling framework for designing robust synthetic controllers.

Future directions

In addition to demonstrating the significant potential of extended S-Systems for plant clock modelling, our work also provides potential insight into the shift from representing protein

pathways with two equations for separate cytoplasmic and nuclear compartments (e.g. JL2005, JL2006), to a single, aggregated equation (e.g. AP2012, KF2014, MF2016) in the established suite of *Arabidopsis* models. Specifically, our sine sweeping results show that in the models employing two pathway equations, the resulting second-order transfer function relating mRNA to active protein can be well-approximated with a first-order transfer function, owing to the two poles being far apart (*i.e.* due to time-scale separation), as shown in Fig 6 and S11 Fig. A single pathway equation would therefore have yielded the same results in each case, implying that in the construction of these models, the data-fitting procedure did not exploit the greater flexibility conferred by the second-order transfer function and/or the increased flexibility was not required to obtain good fits. This apparent redundancy, together with a desire for reduced parametrisations, may partly have driven the subsequent models to jettison separate compartments. It would therefore be of interest to extend our extended S-System framework to integrate distributed delay-based models of protein pathways [19, 72, 73], as these yield tunable transfer functions of arbitrary order, whilst maintaining a compact parametrisation [16]. In addition, although our modification to the original S-System framework extends the range of transcriptional regulation mechanisms that can be modelled beyond the multi-input AND gate of the original formulation (*cf.* Eqs (1) and (2)), it still does not cover the full range of biologically realistic logic gates. This limitation could be addressed by modifying the first term of Eq (2) to encode an S-System implementation of the nested analysing Boolean functions associated with n_i^p inputs [74]. It would be instructive to investigate whether such extensions to the S-System framework resulted in models yielding more accurate fits to data (e.g. the fits of MF2016KS to experimental DD release recordings), but with reduced complexity compared to Michaelis-Menten based models.

Finally, we note that although we have focused here on plant circadian clock models, the approach presented is potentially applicable to all GRNs characterised by entrainable, periodic oscillations.

Supporting information

S1 Text. Supporting Text.

(PDF)

S1 Fig. Plant clock model JL2006S—Optimal fits to synthetic training data. Blue solid lines show timeseries generated by JL2006 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant light (LL). Black dashed lines show timeseries obtained by optimising the parameters of JL2006S to this data in the same simulated light environment (see Fig 2B for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S2 Fig. Plant clock model JL2006S—Fits to synthetic validation data. Blue solid lines show timeseries generated by JL2006 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant dark (DD). Black dashed lines show timeseries generated by JL2006S in the same simulated light environment using the parameters optimised to the training data (see Fig 2C for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S3 Fig. Plant clock model AP2012S—Optimal fits to synthetic training data. Blue solid lines show timeseries generated by AP2012 from its nominal parameter values for a simulated

transition between a 12L:12D light-dark cycle and constant light (LL). Black dashed lines show timeseries obtained by optimising the parameters of AP2012S to this data in the same simulated light environment (see Fig 3B for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S4 Fig. Plant clock model AP2012S—Fits to synthetic validation data. Blue solid lines show timeseries generated by AP2012 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant dark (DD). Black dashed lines show timeseries generated by AP2012S in the same simulated light environment using the parameters optimised to the training data (see Fig 3C for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S5 Fig. Plant clock model KF2014S—Optimal fits to synthetic training data. Blue solid lines show timeseries generated by KF2014 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant light (LL). Black dashed lines show timeseries obtained by optimising the parameters of KF2014S to this data in the same simulated light environment (see Fig 4B for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S6 Fig. Plant clock model KF2014S—Fits to synthetic validation data. Blue solid lines show timeseries generated by KF2014 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant dark (DD). Black dashed lines show timeseries generated by KF2014S in the same simulated light environment using the parameters optimised to the training data (see Fig 4C for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S7 Fig. Plant clock models MF2016KS, MF2016KSorig and MF2016K—Optimal fits to experimental training data. Blue solid lines show timeseries recorded experimentally during a transition between a 12L:12D light-dark cycle and constant light (LL). Black (MF2016KS) and green (MF2016KSorig) dashed lines show timeseries obtained by optimising the parameters of the S-System models to this data in the same simulated light environment. Red dashed lines show optimal fits of MF2016K to the same data, obtained previously in [15] (see Fig 5B for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S8 Fig. Plant clock models MF2016KS, MF2016KSorig and MF2016K—Fits to experimental validation data. Blue solid lines show timeseries recorded experimentally during a transition between a 12L:12D light-dark cycle and constant dark (DD). Black (MF2016KS) and green (MF2016KSorig) dashed lines show timeseries generated by the S-System models in the same simulated light environment using the parameters optimised to the training data. Red dashed lines show the corresponding fits of MF2016K to the same data (see Fig 5C for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

(EPS)

S9 Fig. Correlation method. The output of the sine sweeping test $y(t)$ is correlated with $\sin \omega t$ and $\cos \omega t$ prior to averaging to obtain the corresponding magnitude and phase values required

to construct a Bode plot.
(EPS)

S10 Fig. Linear approximations to nonlinear Y protein degradation in JL2006. Blue lines show how the degradation rates of cytoplasmic Y protein (top panel) and nuclear Y protein (bottom panel) depend on the corresponding expression levels, $c_Y^{(c)}$ and $c_Y^{(n)}$, respectively. In each case, degradation rate is plotted for expression levels ranging between 0 and the maximum level observed in the synthetic training and validation datasets (see S1 and S2 Figs). In these ranges, the nonlinear functions are well-approximated by linear fits (red lines), the gradients of which are taken as the values of γ_{cy} and γ_{nu} used to derive eq. (S3.5) in S1 Text.
(EPS)

S11 Fig. Bode plot relating input Y mRNA to output Y nuclear protein in JL2006. Blue lines represent the second-order system given by eq. (S3.5) in S1 Text. Red lines represent the first-order system given by eq. (S3.7) that approximates eq. (S3.5).
(EPS)

S12 Fig. Variation in optimised parameter values for the extended S-System models. A-D: Fits of JL2005S, JL2006S, AP2012S and KF2014S to synthetic data. **E:** Fits of MF2016KS to experimental data. Boxplots show parameter distributions obtained from six independent optimisation runs. In each boxplot, the horizontal line denotes the median value, the edges of the box are the 25th and 75th percentiles, the whiskers denote the most extreme datapoints not considered to be outliers, and outliers are plotted as red crosses. Model parameter indices are defined in S3 Table (JL2005S), S6 Table (JL2006S), S9 Table (AP2012S), S12 Table (KF2014S) and S14 Table (MF2016KS). In B-E, the thick black horizontal lines separate parameters whose values are plotted with respect to the left and right y-axes.
(EPS)

S1 Table. Variables used in the equations for plant clock models JL2005 [11], JL2006 [43], AP2012 [44], KF2014 [45] and MF2016K [15].
(EPS)

S2 Table. Nominal parameter values for JL2005, which were used to generate synthetic data. The parameters were taken from Figure 5 of [11].
(EPS)

S3 Table. Optimal parameter values $\hat{\Theta}_{LL}^s$ for the extended S-System formulation JL2005S of JL2005, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in S12 Fig.
(EPS)

S4 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting JL2005S to the synthetic training and validation datasets.
(EPS)

S5 Table. Nominal parameter values for JL2006, which were used to generate synthetic data. The parameter values were taken from inline Supplementary Table 1 of [43].
(EPS)

S6 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation JL2006S of JL2006, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in [S12 Fig](#).

(EPS)

S7 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting JL2006S to the synthetic training and validation datasets.

(EPS)

S8 Table. Nominal parameter values for AP2012, used to generate synthetic data. The parameter values were taken from Supplemental Table 1 of [44].

(EPS)

S9 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation AP2012S of AP2012, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in [S12 Fig](#).

(EPS)

S10 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting AP2012S to the synthetic training and validation datasets.

(EPS)

S11 Table. Nominal parameter values for KF2014, which were used to generate synthetic data. The parameter values were taken from Table 3 (Parameter Set 2)[†] and Table 4 in Supporting Information Text S1 of [45] ([†]We note that in the original paper, Table 3 incorrectly lists parameter a_3 as a_1 —this has been fixed in our version of the table).

(EPS)

S12 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation KF2014S of KF2014, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in [S12 Fig](#).

(EPS)

S13 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting KF2014 to the synthetic training and validation datasets.

(EPS)

S14 Table. Optimal parameter values $\hat{\Theta}_{LL}^E$ for the extended S-System formulation MF2016KS of MF2016K, obtained by fitting to the experimental training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with the values obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing,

counting left to right across rows, that is used in [S12 Fig](#).
(EPS)

S15 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting MF2016KS to the experimental training and validation data-sets.

(EPS)

S16 Table. Optimal parameter values $\hat{\Theta}_{LL}^E$ for the original S-System formulation MF2016KSorig of MF2016K, obtained by fitting to the experimental training data.

(EPS)

S17 Table. Optimal parameter values for MF2016K, which were obtained previously in [\[15\]](#) by fitting to the experimental training data used in this study with the same optimisation method. The values are reproduced from Tables S2 and S4 in the Supporting Information of [\[15\]](#).

(EPS)

Acknowledgments

OEA would like to thank Jonathan Fieldsend for useful discussions regarding the parameter variability analysis.

Author Contributions

Conceptualization: Mathias Foo, Declan G. Bates, Ozgur E. Akman.

Formal analysis: Mathias Foo, Ozgur E. Akman.

Funding acquisition: Declan G. Bates.

Investigation: Mathias Foo, Ozgur E. Akman.

Methodology: Mathias Foo, Ozgur E. Akman.

Project administration: Ozgur E. Akman.

Software: Mathias Foo.

Supervision: Declan G. Bates, Ozgur E. Akman.

Validation: Mathias Foo.

Writing – original draft: Mathias Foo, Ozgur E. Akman.

Writing – review & editing: Mathias Foo, Declan G. Bates, Ozgur E. Akman.

References

1. Nagel DH, Kay SA. Complexity in the wiring and regulation of plant circadian networks. *Curr Biol*. 2012; 22(16):R648–R657. <https://doi.org/10.1016/j.cub.2012.07.025> PMID: 22917516
2. Broody S, Harris S. Circadian rhythms in *Neurospora*: spatial differences in pyridine nucleotide levels. *Science*. 1973; 180:498–500. <https://doi.org/10.1126/science.180.4085.498>
3. Sehgal A, Price J, Man B, Young M. Loss of circadian behavioral rhythms and per RNA oscillations in the *Drosophila* mutant timeless. *Science*. 1994; 263:1603–1606.
4. McClung R, Harris S. Plant circadian rhythms. *Plant Cell*. 2006; 18:792–803. <https://doi.org/10.1105/tpc.106.040980> PMID: 16595397

5. Lange T, Dimitrov S, Fehm HL, Westermann J, Born J. Shift of monocyte function toward cellular immunity during sleep. *Arch Intern Med*. 2006; 166:1695–1700. <https://doi.org/10.1001/archinte.166.16.1695> PMID: 16983046
6. Ruger M, Scheer FAJL. Effects of circadian disruption on the cardiometabolic system. *Rev Endocr Meta Disord*. 2009; 10:245–260. <https://doi.org/10.1007/s11154-009-9122-8>
7. Khapre RV, Samsa WE, Kondatov RV. Circadian regulation of cell cycle: molecular connections between aging and the circadian clock. *Ann Med*. 2010; 42:1695–1700. <https://doi.org/10.3109/07853890.2010.499134>
8. Westrich L, Sprouse J. Circadian rhythm dysregulation in bipolar disorder. *Curr Opin Invest Drugs*. 2010; 11:779–787.
9. Callaway E, Ledford H. Medicine Nobel awarded for work on circadian clocks. *Nature*. 2017; 550:18. <https://doi.org/10.1038/nature.2017.22736> PMID: 28980662
10. Ibanez C. Scientific background discoveries of molecular mechanisms controlling the circadian rhythm. The Nobel Assembly at Karolinska Institutet; 2017. Available from: <https://www.nobelprize.org/uploads/2018/06/advanced-medicineprize2017.pdf>.
11. Locke JCW, Millar AJ, Turner MS. Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *J Theor Biol*. 2005; 234:383–393. <https://doi.org/10.1016/j.jtbi.2004.11.038> PMID: 15784272
12. Akman OE, Watterson S, Parton A, Binns N, Millar AJ, Ghazal P. Digital clocks: simple Boolean models can qualitatively describe circadian systems. *J R Soc Interface*. 2012; 9:2365–2382. <https://doi.org/10.1098/rsif.2012.0080> PMID: 22499125
13. Adams R, Clark A, Yamaguchi A, Hanlon N, Tsorman N, Ali S, et al. SBSI: an extensible distributed software infrastructure for parameter estimation in systems biology. *Bioinformatics*. 2013; 29(5):664–665. <https://doi.org/10.1093/bioinformatics/btt023> PMID: 23329415
14. DeCaluwe J, Xiao Q, Hermans C, Verbruggen N, Leloup JL, Gonze D. A compact model for the complex plant circadian clock. *Front Plant Sci*. 2016; 7(74).
15. Foo M, Somers DE, Kim PJ. Kernel architecture of the genetic circuitry of the *Arabidopsis* circadian system. *PLoS Comput Biol*. 2016; 12(2):e1004748. <https://doi.org/10.1371/journal.pcbi.1004748> PMID: 26828650
16. Tokuda IT, Akman OE, Locke JCW. Reducing the complexity of mathematical models for the plant circadian clock by distributed delays. *J Theor Biol*. 2019; 463:155–166. <https://doi.org/10.1016/j.jtbi.2018.12.014> PMID: 30550861
17. Tyson JJ, Hong CI, Thron CD, Novak B. A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. *Biophys J*. 1999; 77:2411–2417. [https://doi.org/10.1016/S0006-3495\(99\)77078-5](https://doi.org/10.1016/S0006-3495(99)77078-5) PMID: 20540926
18. Leloup JC, Goldbeter A. Toward a detailed computational model for the mammalian circadian clock. *Proc Natl Acad Sci USA*. 2003; 100:7051–7056. <https://doi.org/10.1073/pnas.1132112100> PMID: 12775757
19. Akman OE, Rand DA, Brown PE, Millar AJ. Robustness from flexibility in the fungal circadian clock. *BMC Syst Biol*. 2010; 4(1):88. <https://doi.org/10.1186/1752-0509-4-88> PMID: 20576110
20. Savageau MA. Biochemical systems analysis II. The steady state solutions for an n-pool system using a power-law approximation. *J Theor Biol*. 1969; 25:370–379. [https://doi.org/10.1016/s0022-5193\(69\)80027-5](https://doi.org/10.1016/s0022-5193(69)80027-5) PMID: 5387047
21. Kikuchi S, Tominaga D, Arita M, Takahashi K, Tomita M. Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*. 2003; 19(5):643–650. <https://doi.org/10.1093/bioinformatics/btg027> PMID: 12651723
22. Kimura S, Ide K, Kashihara A, Kano M, Hatakeyama M, Masui R, et al. Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*. 2005; 21(7):1154–1163. <https://doi.org/10.1093/bioinformatics/bti071> PMID: 15514004
23. Savageau MA. Design principles for elementary gene circuits: elements, methods, and examples. *Chaos*. 2001; 11:142–159. <https://doi.org/10.1063/1.1349892> PMID: 12779449
24. Voit E, Martens HA, Omholt SW. 150 years of mass action law. *PLoS Comput Biol*. 2015; 11:e1004012. <https://doi.org/10.1371/journal.pcbi.1004012> PMID: 25569257
25. Lin CL, Liu YW, Chuang CH. Analysis of signal transduction networks in Michaelis-Menten equations and S-systems. *Int J Biol Biomed Eng*. 2008; 2:69–78.
26. Rizk ML, Liao JC. Ensemble modeling and related mathematical modeling of metabolic networks. *J Taiwan Inst Chem E*. 2009; 40:595–601. <https://doi.org/10.1016/j.jtice.2009.05.003>

27. Sorribas A, Savageau MA. A comparison of variant theories of intact biochemical systems. I. enzyme-enzyme interactions and biochemical systems theory. *Math Biosci.* 1989; 94:161–193. [https://doi.org/10.1016/0025-5564\(89\)90064-3](https://doi.org/10.1016/0025-5564(89)90064-3) PMID: 2520168
28. Sorribas A, Savageau MA. A comparison of variant theories of intact biochemical systems II: flux-oriented and metabolic control theories. *Math Biosci.* 1989; 94:195–238. [https://doi.org/10.1016/0025-5564\(89\)90065-5](https://doi.org/10.1016/0025-5564(89)90065-5) PMID: 2520169
29. Sorribas A, Savageau MA. Strategies for representing metabolic pathways within biochemical systems theory: Reversible pathways. *Math Biosci.* 1989; 94:239–269. [https://doi.org/10.1016/0025-5564\(89\)90066-7](https://doi.org/10.1016/0025-5564(89)90066-7) PMID: 2520170
30. Voit EO, Savageau M. Accuracy of alternative representations for integrated biochemical systems. *Biochemistry.* 1987; 26(21):6869–6880. <https://doi.org/10.1021/bi00395a042> PMID: 3427048
31. Curto R, Sorribas A, Cascante M. Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: model definition and nomenclature. *Math Biosci.* 1995; 130:25–50. [https://doi.org/10.1016/0025-5564\(94\)00092-e](https://doi.org/10.1016/0025-5564(94)00092-e) PMID: 7579901
32. Curto R, Voit EO, Sorribas A, Cascante M. Mathematical models of purine metabolism in man. *Math Biosci.* 1998; 151:1–49. [https://doi.org/10.1016/s0025-5564\(98\)10001-9](https://doi.org/10.1016/s0025-5564(98)10001-9) PMID: 9664759
33. Alvarez-Vasquez F, Sims KJ, Hannun YA, Voit EO. Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models. *J Theor Biol.* 2004; 226:265–291. <https://doi.org/10.1016/j.jtbi.2003.08.010> PMID: 14643642
34. Wang CCN, Chang PC, Ng KL, Chang CM, Sheu PCY, Tsai JJP. A model comparison study of the flowering time regulatory network in *Arabidopsis*. *BMC Syst Biol.* 2014; 8. <https://doi.org/10.1186/1752-0509-8-15>
35. Savageau M. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology.* Reading, MA: Addison-Wesley; 1976.
36. Hlavacek WS, Savageau MA. Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol.* 1996; 255:121–139. <https://doi.org/10.1006/jmbi.1996.0011> PMID: 8568860
37. Alves R, Savageau MA. Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: Differences in molecular structure and physiological function. *Mol Microbiol.* 2003; 48:25–51. <https://doi.org/10.1046/j.1365-2958.2003.03344.x> PMID: 12657043
38. Kurosawa G, Mochizuki A, Iwasa Y. Comparative study of circadian clock models, in search of processes promoting oscillation. *J Theor Biol.* 2002; 216(2):193–208. <https://doi.org/10.1006/jtbi.2002.2546> PMID: 12079371
39. Voit EO. *Computational analysis of biochemical systems. A practical guide for biochemists and molecular biologists.* Cambridge University Press, Cambridge, IUK; 2000.
40. Voit E. Biochemical systems theory: a review. *ISRN Biomath.* 2013; p. 1–53. <https://doi.org/10.1155/2013/897658>
41. Alon U. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall, London; 2006.
42. Buchler N, Gerland U, Hwa T. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci USA.* 2003; 100(9):5136–5141. <https://doi.org/10.1073/pnas.0930314100> PMID: 12702751
43. Locke JCW, Kozma-Bognar L, Gould PD, Feher B, Kevei E, Nagy F, et al. Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Mol Syst Biol.* 2006; 2(59). <https://doi.org/10.1038/msb4100102> PMID: 17102804
44. Pokhilko A, Fernandez AP, Edwards KD, Southern MM, Halliday KJ, Millar AJ. The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol Syst Biol.* 2012; 8(574). <https://doi.org/10.1038/msb.2012.6> PMID: 22395476
45. Fogelmark K, Troein C. Rethinking transcriptional activation in the *Arabidopsis* circadian clock. *PLoS Comput Biol.* 2014; 10(7):e1003705. <https://doi.org/10.1371/journal.pcbi.1003705> PMID: 25033214
46. Lagarias J, Reeds J, Wright M, Wright P. Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM J Optim.* 1998; 9(1):112–147. <https://doi.org/10.1137/S1052623496303470>
47. Burnham KP, Anderson DR. *Information and Likelihood Theory: A Practical Information-Theoretic Approach.* Springer-Verlag, New York; 2002.
48. Burnham KP, Anderson DR. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol Methods Res.* 2004; 33(2):261–304. <https://doi.org/10.1177/0049124104268644>
49. Banks HT, Joyner ML. AIC under the framework of least squares estimation. *Appl Math Lett.* 2017; 74:33–45. <https://doi.org/10.1016/j.aml.2017.05.005>

50. Wagenmakers EJ, Farrell S. AIC model selection using Akaike weights. *Psychon Bull Rev.* 2004; 11(1):192–196. <https://doi.org/10.3758/bf03206482> PMID: 15117008
51. Huang W, Perez-Garcia P, Pokhilko A, Millar AJ, Antoshekin I, Riechmann JL, et al. Mapping the core of the *Arabidopsis* circadian clock defines the network structure of the oscillator. *Science.* 2012; 336:75–79. <https://doi.org/10.1126/science.1219075> PMID: 22403178
52. Gendron JM, Pruneda-Paz JL, Doherty CJ, Gross AM, Kang SE, Kay SA. *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc Natl Acad Sci USA.* 2012; 109:3167–3172. <https://doi.org/10.1073/pnas.1200355109> PMID: 22315425
53. Yu JW, Rubio V, Lee NY, Bai S, Lee SY, Kim SS, et al. COP1 and ELF3 control circadian function and photoperiodic flowering by regulating GI stability. *Mol Cell.* 2008; 32:617–630. <https://doi.org/10.1016/j.molcel.2008.09.026> PMID: 19061637
54. Wang L, Fujiwara S, Somers DE. PRR5 regulates phosphorylation, nuclear import and subnuclear localization of TOC1 in the *Arabidopsis* circadian clock. *EMBO J.* 2010; 29:1903–1915. <https://doi.org/10.1038/emboj.2010.76> PMID: 20407420
55. Ljung L. *System Identification: Theory for the User.* 2nd ed. Prentice Hall; 1999.
56. Nise N. *Control Systems Engineering.* 5th ed. John Wiley and Sons, Inc.; 2008.
57. Mizoguchi T, Wheatley K, Hanzawa Y, Mizoguchi M, Song HR, Carre IA, et al. *LHY* and *CCA1* are partially redundant genes required to maintain circadian rhythms in *Arabidopsis*. *Dev Cell.* 2002; 2:629–641. [https://doi.org/10.1016/S1534-5807\(02\)00170-3](https://doi.org/10.1016/S1534-5807(02)00170-3)
58. Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, et al. Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science.* 2000; 289:768–771. <https://doi.org/10.1126/science.289.5480.768> PMID: 10926537
59. Farre EM, Harmer SL, Harmon FG, Yanovsky MJ, Kay SA. Overlapping and distinct roles of *PRR7* and *PRR9* in the *Arabidopsis* circadian clock. *Curr Biol.* 2005; 15:47–54. <https://doi.org/10.1016/j.cub.2004.12.067> PMID: 15649364
60. Somers DE, Schultz TF, Milnamow M, Kay SA. *ZEITLUPE* encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell.* 2000; 101:319–329. [https://doi.org/10.1016/s0092-8674\(00\)80841-7](https://doi.org/10.1016/s0092-8674(00)80841-7) PMID: 10847686
61. Gould PD, Locke JCW, Larue C, Southern MM, Davis SJ, Hanano S, et al. The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. *Plant Cell.* 2006; 18:1177–1187. <https://doi.org/10.1105/tpc.105.039990> PMID: 16617099
62. McWatters HG, Kolmos E, Hall A, Doyle MR, Amasino RM, Gyula P, et al. *ELF4* is required for oscillatory properties of the circadian clock. *Plant Physiol.* 2007; 144:391–401. <https://doi.org/10.1104/pp.107.096206> PMID: 17384164
63. McWatters HG, Bastow RM, Hall A, Millar AJ. The *ELF3 zeitnehmer* regulates light signalling to the circadian clock. *Nature.* 2000; 408:716–720. <https://doi.org/10.1038/35047079> PMID: 11130072
64. Onai K, Ishiura M. *PHYTOCLOCK 1* encoding a novel GARP protein essential for the *Arabidopsis* circadian clock. *Genes Cells.* 2005; 10:963–972. <https://doi.org/10.1111/j.1365-2443.2005.00892.x> PMID: 16164597
65. Herero E, Kolmos E, Bujdoso N, Yuan Y, Wang M, Berns MC, et al. *EARLY FLOWERING4* recruitment of *EARLY FLOWERING3* in the nucleus sustains the *Arabidopsis* circadian clock. *Plant Cell.* 2012; 24:428–443. <https://doi.org/10.1105/tpc.111.093807>
66. Martin-Tryon EL, Kreps JA, Harmer SL. *GIGANTEA* acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation. *Plant Physiol.* 2007; 143:473–486. <https://doi.org/10.1104/pp.106.088757> PMID: 17098855
67. Doyle MR, Davis SJ, Bastow RM, McWatters HG, Kozma-Bognar L, Nagy F, et al. The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature.* 2002; 419:74–77. <https://doi.org/10.1038/nature00954> PMID: 12214234
68. Kevei E, Gyula P, Hall A, Kozma-Bognar L, Kim WY, Eriksson ME, et al. Forward genetic analysis of the circadian clock separates the multiple functions of *ZEITLUPE*. *Plant Physiol.* 2006; 140:933–945. <https://doi.org/10.1104/pp.105.074864> PMID: 16428597
69. Adams S, Manfield I, Stockley P, Carre IA. Revised morning loops of the *Arabidopsis* circadian clock based on analyses of direct regulatory interactions. *PLoS One.* 2015; 10:e0143943. <https://doi.org/10.1371/journal.pone.0143943> PMID: 26625126
70. Foo M, Kim J, Bates DG. Modelling and control of gene regulatory networks for perturbation mitigation. *IEEE/ACM Trans Comput Biol Bioinform.* 2018; 16:583–595. <https://doi.org/10.1109/TCBB.2017.2771775> PMID: 29994499
71. Srinath S, Gunawan R. Parameter identifiability of power-law biochemical system models. *J Biotechnol.* 2010; 149:132–140. <https://doi.org/10.1016/j.jbiotec.2010.02.019> PMID: 20197073

72. Akman OE, Locke JCW, Tang S, Carré I, Millar AJ, Rand DA. Isoform switching facilitates period control in the *Neurospora crassa* circadian clock. *Mol Syst Biol*. 2008; 4:64. <https://doi.org/10.1038/msb.2008.5>
73. MacDonald N. Biological delay systems: linear stability theory. vol. 9. Cambridge University Press; 2008.
74. Nikolajewa S, Friedel M, Wilhelm T. Boolean networks with biologically relevant rules show ordered behavior. *Biosystems*. 2007; 90(1):40–47. <https://doi.org/10.1016/j.biosystems.2006.06.007> PMID: [17188807](https://pubmed.ncbi.nlm.nih.gov/17188807/)

A simplified modelling framework facilitates more complex representations of plant circadian clocks

Mathias Foo¹, Declan G. Bates², Ozgur E. Akman^{3*},

1 School of Mechanical, Aerospace and Automotive Engineering, Coventry University, Coventry, CV1 5FB, UK

2 Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, CV4 7AL, UK

3 College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK

* O.E.Akman@exeter.ac.uk

Supporting Text

Contents

1	Model equations	1
1.1	JL2005: original equations	2
1.2	JL2005S: extended S-System formulation	2
1.3	JL2006: original equations	2
1.4	JL2006S: extended S-System formulation	3
1.5	AP2012: original equations	4
1.6	AP2012S: extended S-System formulation	6
1.7	KF2014: original equations	7
1.8	KF2014S: extended S-System formulation	9
1.9	MF2016K: original equations	11
1.10	MF2016KS: extended S-System formulation	12
1.11	MF2016KSortig: original S-System formulation	13
2	The sine-sweeping method	14
3	Transfer function analysis of protein shuttling in JL2006	15
	References	16

1 Model equations

In each model below, $c_i^{(m)}(t)$ and $c_i(t)$ denote the cellular concentration of the mRNA and protein of gene i , respectively. For the models where cyttoplasmic and nuclear protein are treated separately, these two forms are labelled as $c_i^{(c)}(t)$ and $c_i^{(n)}(t)$, respectively. The different gene products labelled by i are listed in **S1 Table**. $c_P(t)$ (or $c_P^{(n)}(t)$) denotes the light-activated PIF3-like protein introduced originally in JL2005 [S1] and subsequently used in all the other clock models considered in this study. The light input, $L_I(t)$, is modelled as a periodic square wave that switches between 0 and 1 (see eq. (3) of the main paper).

1.1 JL2005: original equations

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= q_1 c_P^{(n)} L_I(t) + \frac{n_1 c_T^{(n)a}}{g_1^a + c_T^{(n)a}} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}}, \\
\frac{dc_L^{(c)}}{dt} &= p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}}, \\
\frac{dc_L^{(n)}}{dt} &= r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}}, \\
\frac{dc_T^{(m)}}{dt} &= \frac{n_2 g_2^b}{g_2^b + c_L^{(n)b}} - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}}, \\
\frac{dc_T^{(c)}}{dt} &= p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - \frac{m_5 c_T^{(c)}}{k_5 + c_T^{(c)}}, \\
\frac{dc_T^{(n)}}{dt} &= r_3 c_T^{(c)} - r_4 c_T^{(n)} - \frac{m_6 c_T^{(n)}}{k_6 + c_T^{(n)}}, \\
\frac{dc_P^{(n)}}{dt} &= p_3(1 - L_I(t)) - \frac{m_7 c_P^{(n)}}{k_7 + c_P^{(n)}} - q_2 c_P^{(n)} L_I(t).
\end{aligned} \tag{S1.1}$$

1.2 JL2005S: extended S-System formulation

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= \alpha_1 c_T^{(n)g_{1,1}} - \beta_{1,1} c_L^{(m)} + \gamma_{1,1} c_P^{(n)} L_I(t), \\
\frac{dc_L^{(c)}}{dt} &= \alpha_2 c_L^{(m)g_{2,1}} - \beta_{2,1} c_L^{(c)}, \\
\frac{dc_L^{(n)}}{dt} &= \alpha_3 c_L^{(c)g_{3,1}} - \beta_{3,1} c_L^{(n)}, \\
\frac{dc_T^{(m)}}{dt} &= \alpha_4 c_L^{(n)g_{4,1}} - \beta_{4,1} c_T^{(m)}, \\
\frac{dc_T^{(c)}}{dt} &= \alpha_5 c_T^{(m)g_{5,1}} - \beta_{5,1} c_T^{(c)}, \\
\frac{dc_T^{(n)}}{dt} &= \alpha_6 c_T^{(c)g_{6,1}} - \beta_{6,1} c_T^{(n)}, \\
\frac{dc_P^{(n)}}{dt} &= -\beta_{7,1} c_P^{(n)} + \gamma_{7,1}(1 - L_I(t)) + \gamma_{7,2} c_P^{(n)} L_I(t).
\end{aligned} \tag{S1.2}$$

1.3 JL2006: original equations

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= \frac{g_0^\alpha}{g_0^\alpha + c_A^{(n)\alpha}} \cdot \left(\left(q_1 c_P^{(n)} + n_0 \right) L_I(t) + \frac{n_1 c_X^{(n)a}}{g_1^a + c_X^{(n)a}} \right) - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}}, \\
\frac{dc_L^{(c)}}{dt} &= p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}}, \\
\frac{dc_L^{(n)}}{dt} &= r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_T^{(m)}}{dt} &= \frac{n_2 c_Y^{(n)b}}{g_2^b + c_Y^{(n)b}} \cdot \frac{g_3^c}{g_3^c + c_L^{(n)c}} - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}}, \\
\frac{dc_T^{(c)}}{dt} &= p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - (m_5(1 - L_I(t)) + m_6) \frac{c_T^{(c)}}{k_5 + c_T^{(c)}}, \\
\frac{dc_T^{(n)}}{dt} &= r_3 c_T^{(c)} - r_4 c_T^{(n)} - (m_7(1 - L_I(t)) + m_8) \frac{c_T^{(n)}}{k_6 + c_T^{(n)}}, \\
\frac{dc_X^{(m)}}{dt} &= \frac{n_3 c_T^{(n)d}}{g_4^d + c_T^{(n)d}} - \frac{m_9 c_X^{(m)}}{k_7 + c_X^{(m)}}, \\
\frac{dc_X^{(c)}}{dt} &= p_3 c_X^{(m)} - r_5 c_X^{(c)} + r_6 c_X^{(n)} - \frac{m_{10} c_X^{(c)}}{k_8 + c_X^{(c)}}, \\
\frac{dc_X^{(n)}}{dt} &= r_5 c_X^{(c)} - r_6 c_X^{(n)} - \frac{m_{11} c_X^{(n)}}{k_9 + c_X^{(n)}}, \\
\frac{dc_Y^{(m)}}{dt} &= \frac{g_6^f}{g_6^f + c_L^{(n)f}} \cdot \left(q_2 c_P^{(n)} L_I(t) + \frac{(n_4 L_I(t) + n_5) g_5^e}{g_5^e + c_T^{(n)e}} \right) - \frac{m_{12} c_Y^{(m)}}{k_{10} + c_Y^{(m)}}, \\
\frac{dc_Y^{(c)}}{dt} &= p_4 c_Y^{(m)} - r_7 c_Y^{(c)} + r_8 c_Y^{(n)} - \frac{m_{13} c_Y^{(c)}}{k_{11} + c_Y^{(c)}}, \\
\frac{dc_Y^{(n)}}{dt} &= r_7 c_Y^{(c)} - r_8 c_Y^{(n)} - \frac{m_{14} c_Y^{(n)}}{k_{12} + c_Y^{(n)}}, \\
\frac{dc_P^{(n)}}{dt} &= p_5(1 - L_I(t)) - \frac{m_{15} c_P^{(n)}}{k_{13} + c_P^{(n)}} - q_3 c_P^{(n)} L_I(t), \\
\frac{dc_A^{(m)}}{dt} &= q_4 c_P^{(n)} L_I(t) + \frac{n_6 c_L^{(n)g}}{g_7^g + c_L^{(n)g}} - \frac{m_{16} c_A^{(m)}}{k_{14} + c_A^{(m)}}, \\
\frac{dc_A^{(c)}}{dt} &= p_6 c_A^{(m)} - r_9 c_A^{(c)} + r_{10} c_A^{(n)} - \frac{m_{17} c_A^{(c)}}{k_{15} + c_A^{(c)}}, \\
\frac{dc_A^{(n)}}{dt} &= r_9 c_A^{(c)} - r_{10} c_A^{(n)} - \frac{m_{18} c_A^{(n)}}{k_{16} + c_A^{(n)}}. \tag{S1.3}
\end{aligned}$$

1.4 JL2006S: extended S-System formulation

$$\begin{aligned}
\frac{dc_P^{(n)}}{dt} &= -\beta_{1,1} c_P^{(n)} + \gamma_{1,1}(1 - L_I(t)) + \gamma_{1,2} c_P^{(n)} L_I(t), \\
\frac{dc_L^{(m)}}{dt} &= \alpha_2 c_X^{(n)g_{2,1}} c_A^{(n)g_{2,2}} - \beta_{2,1} c_L^{(m)} + \gamma_{2,1} c_P^{(n)} c_A^{(n)g_{2,3}} L_I(t), \\
\frac{dc_L^{(c)}}{dt} &= \alpha_3 c_L^{(m)g_{3,1}} - \beta_{3,1} c_L^{(c)}, \\
\frac{dc_L^{(n)}}{dt} &= \alpha_4 c_L^{(c)g_{4,1}} - \beta_{4,1} c_L^{(n)}, \\
\frac{dc_T^{(m)}}{dt} &= \alpha_5 c_L^{(n)g_{5,1}} c_Y^{(n)g_{5,2}} - \beta_{5,1} c_T^{(m)},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_T^{(c)}}{dt} &= \alpha_6 c_T^{(m)g_{6,1}} - \beta_{6,1} c_T^{(c)} + \gamma_{6,1} c_T^{(c)} (1 - L_I(t)), \\
\frac{dc_T^{(n)}}{dt} &= \alpha_7 c_T^{(c)g_{7,1}} - \beta_{7,1} c_T^{(n)} + \gamma_{7,1} c_T^{(n)} (1 - L_I(t)), \\
\frac{dc_X^{(m)}}{dt} &= \alpha_8 c_T^{(n)g_{8,1}} - \beta_{8,1} c_X^{(m)}, \\
\frac{dc_X^{(c)}}{dt} &= \alpha_9 c_X^{(m)g_{9,1}} - \beta_{9,1} c_X^{(c)}, \\
\frac{dc_X^{(n)}}{dt} &= \alpha_{10} c_X^{(c)g_{10,1}} - \beta_{10,1} c_X^{(n)}, \\
\frac{dc_Y^{(m)}}{dt} &= \alpha_{11} c_L^{(n)g_{11,1}} c_T^{(n)g_{11,2}} - \beta_{11,1} c_Y^{(m)} + \gamma_{11,1} c_P^{(n)} c_L^{(n)g_{11,3}} L_I(t) + \gamma_{11,2} c_L^{(n)g_{11,4}} c_T^{(n)g_{11,5}} L_I(t), \\
\frac{dc_Y^{(c)}}{dt} &= \alpha_{12} c_Y^{(m)g_{12,1}} - \beta_{12,1} c_Y^{(c)}, \\
\frac{dc_Y^{(n)}}{dt} &= \alpha_{13} c_Y^{(c)g_{13,1}} - \beta_{13,1} c_Y^{(n)}, \\
\frac{dc_A^{(m)}}{dt} &= \alpha_{14} c_L^{(n)g_{14,1}} - \beta_{14,1} c_A^{(m)} + \gamma_{14,1} c_P^{(n)} L_I(t), \\
\frac{dc_A^{(c)}}{dt} &= \alpha_{15} c_A^{(m)g_{15,1}} - \beta_{15,1} c_A^{(c)}, \\
\frac{dc_A^{(n)}}{dt} &= \alpha_{16} c_A^{(c)g_{16,1}} - \beta_{16,1} c_A^{(n)}. \tag{S1.4}
\end{aligned}$$

1.5 AP2012: original equations

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= q_1 L_I(t) c_P + \frac{n_1 g_1^a}{g_1^a + (c_{P9} + c_{P7} + c_{NI} + c_T)^a} - (m_1 L_I(t) + m_2 (1 - L_I(t))) c_L^{(m)}, \\
\frac{dc_L}{dt} &= (p_1 L_I(t) + p_2) c_L^{(m)} - m_3 c_L - \frac{p_3 c_L^c}{g_2^c + c_L^c}, \\
\frac{dc_{Lmod}}{dt} &= \frac{p_3 c_L^c}{g_3^c + c_L^c} - m_4 c_{Lmod}, \\
\frac{dc_P}{dt} &= p_7 (1 - L_I(t)) (1 - c_P) - m_{11} c_P L_I(t), \\
\frac{dc_{P9}^{(m)}}{dt} &= q_3 L_I(t) c_P + \frac{g_8}{g_8 + c_{EC}} \left(n_4 + \frac{n_7 c_L^e}{g_9^e + c_L^e} \right) - m_{12} c_{P9}^{(m)}, \\
\frac{dc_{P9}}{dt} &= p_8 c_{P9}^{(m)} - (m_{13} + m_{22} (1 - L_I(t))) c_{P9}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \frac{n_8 c_{Ltot}^e}{g_{10}^e + c_{Ltot}^e} + \frac{n_9 c_{P9}^f}{g_{11}^f + c_{P9}^f} - m_{14} c_{P7}^{(m)}, \\
\frac{dc_{P7}}{dt} &= p_9 c_{P7}^{(m)} - (m_{15} + m_{23} (1 - L_I(t))) c_{P7}, \\
\frac{dc_{NI}^{(m)}}{dt} &= \frac{n_{10} c_{Lmod}^e}{g_{12}^e + c_{Lmod}^e} + \frac{n_{11} c_{P7}^b}{g_{13}^b + c_{P7}^b} - m_{16} c_{NI}^{(m)}, \\
\frac{dc_{NI}}{dt} &= p_{10} c_{NI}^{(m)} - (m_{17} + m_{24} (1 - L_I(t))) c_{NI},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_T^{(m)}}{dt} &= n_2 \frac{g_4}{g_4 + c_{EC}} \cdot \frac{g_5^e}{g_5^e + c_L^e} - m_5 c_T^{(m)}, \\
\frac{dc_T}{dt} &= p_4 c_T^{(m)} - (m_6 + m_7(1 - L_I(t))) c_T (p_5 c_{ZTL} + c_{ZG}) - m_8 c_T, \\
\frac{dc_{E4}^{(m)}}{dt} &= n_{13} \frac{g_2}{g_2 + c_{EC}} \cdot \frac{g_6^e}{g_6^e + c_L^e} - m_{34} c_{E4}^{(m)}, \\
\frac{dc_{E4}}{dt} &= p_{23} c_{E4}^{(m)} - m_{35} c_{E4} - p_{25} c_{E4} c_{E3n} + p_{21} c_{E34}, \\
\frac{dc_{E3}^{(m)}}{dt} &= n_3 \frac{g_{16}^e}{g_{16}^e + c_L^e} - m_{26} c_{E3}^{(m)}, \\
\frac{dc_{E3c}}{dt} &= p_{16} c_{E3}^{(m)} - m_9 c_{E3c} c_{COP1c} - p_{17} c_{E3c} c_{Gc} - p_{19} c_{E3c} + p_{20} c_{E3n}, \\
\frac{dc_{E3n}}{dt} &= p_{19} c_{E3c} - p_{20} c_{E3n} - p_{17} c_{E3n} c_{Gn} - m_{30} c_{E3n} c_{COP1d} - m_{29} c_{E3n} c_{COP1n} \\
&\quad + p_{21} c_{E34} - p_{25} c_{E4} c_{E3n}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= n_{13} \frac{g_2}{g_2 + c_{EC}} \cdot \frac{g_6^e}{g_6^e + c_L^e} - m_{34} c_{LUX}^{(m)}, \\
\frac{dc_{LUX}}{dt} &= p_{27} c_{LUX}^{(m)} - m_{39} c_{LUX} - p_{26} c_{LUX} c_{E34}, \\
\frac{dc_{COP1c}}{dt} &= n_5 - p_6 c_{COP1c} - m_{27} c_{COP1c} (1 + p_{15} L_I(t)), \\
\frac{dc_{COP1n}}{dt} &= p_6 c_{COP1c} - n_6 L_I(t) c_{PCOP1n} - n_{14} c_{COP1n} - m_{27} c_{COP1n} (1 + p_{15} L_I(t)), \\
\frac{dc_{COP1d}}{dt} &= n_{14} c_{COP1n} + n_6 L_I(t) c_{PCOP1n} - m_{31} (1 + m_{33} (1 - L_I(t))) c_{COP1d}, \\
\frac{dc_{EGc}}{dt} &= p_{17} c_{E3c} c_{Gc} - m_9 c_{EGc} c_{COP1c} - p_{18} c_{EGc} + p_{31} c_{EGn}, \\
\frac{dc_{EC}}{dt} &= p_{26} c_{LUX} c_{E34} - m_{36} c_{EC} c_{COP1n} - m_{37} c_{EC} c_{COP1d} \\
&\quad - m_{32} c_{EC} \left(1 + p_{24} L_I(t) \frac{c_{Gn.tot}^d}{g_7^d + c_{Gn.tot}^d} \right), \\
\frac{dc_{ZTL}}{dt} &= p_{14} - p_{12} L_I(t) c_{ZTL} c_{Gc} + p_{13} c_{ZG} (1 - L_I(t)) - m_{20} c_{ZTL}, \\
\frac{dc_{ZG}}{dt} &= p_{12} L_I(t) c_{ZTL} c_{Gc} - p_{13} c_{ZG} (1 - L_I(t)) - m_{21} c_{ZG}, \\
\frac{dc_G^{(m)}}{dt} &= q_2 L_I(t) c_P + n_{12} \frac{g_{14}}{g_{14} + c_{EC}} \cdot \frac{g_{15}^e}{g_{15}^e + c_L^e} - m_{18} c_G^{(m)}, \\
\frac{dc_{Gc}}{dt} &= p_{11} c_G^{(m)} - p_{12} L_I(t) c_{ZTL} c_{Gc} + p_{13} c_{ZG} (1 - L_I(t)) - m_{19} c_{Gc} - p_{17} c_{E3c} c_{Gc} \\
&\quad - p_{28} c_{Gc} + p_{29} c_{Gn}, \\
c_{Ltot} &= c_L + c_{Lmod}, \\
c_{E34} &= \frac{p_{25} c_{E4} c_{E3n}}{p_{26} c_{LUX} + p_{21} + m_{37} c_{COP1d} + m_{36} c_{COP1n}}, \\
c_{EGn} &= \frac{p_{18} c_{EGc} + p_{17} c_{E3n} c_{Gn}}{m_9 c_{COP1n} + m_{10} c_{COP1d} + p_{31}}, \\
c_{Gn} &= \frac{p_{28} c_{Gc}}{p_{29} + m_{19} + p_{17} c_{E3n}},
\end{aligned}$$

$$c_{Gn_tot} = c_{Gn} + c_{EGn}. \quad (S1.5)$$

1.6 AP2012S: extended S-System formulation

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= \alpha_1(c_{P9} + c_{P7} + c_{NI} + c_T)^{g_{1,1}} - \beta_{1,1}c_L^{(m)} + \gamma_{1,1}c_PL_I(t) + \gamma_{1,2}c_L^{(m)}L_I(t), \\
\frac{dc_L}{dt} &= \alpha_2c_L^{(m)} - \beta_{2,1}c_L + \gamma_{2,1}c_L^{(m)}L_I(t), \\
\frac{dc_{Lmod}}{dt} &= \alpha_3c_L^{g_{3,1}} - \beta_{3,1}c_{Lmod}, \\
\frac{dc_P}{dt} &= \gamma_{4,1}(1 - L_I(t)) + \gamma_{4,2}c_P(1 - L_I(t)) + \gamma_{4,3}c_PL_I(t), \\
\frac{dc_{P9}^{(m)}}{dt} &= \alpha_5c_L^{g_{5,1}}c_{EC}^{g_{5,2}} - \beta_{5,1}c_{P9}^{(m)} + \gamma_{5,1}L_I(t)c_P, \\
\frac{dc_{P9}}{dt} &= \alpha_6c_{P9}^{(m)} - \beta_{6,1}c_{P9} + \gamma_{6,1}c_{P9}(1 - L_I(t)), \\
\frac{dc_{P7}^{(m)}}{dt} &= \alpha_7(c_L + c_{Lmod})^{g_{7,1}}c_{P9}^{g_{7,2}} - \beta_{7,1}c_{P7}^{(m)}, \\
\frac{dc_{P7}}{dt} &= \alpha_8c_{P7}^{(m)} - \beta_{8,1}c_{P7} + \gamma_{8,1}c_{P7}(1 - L_I(t)), \\
\frac{dc_{NI}^{(m)}}{dt} &= \alpha_9c_{Lmod}^{g_{9,1}}c_{P7}^{g_{9,2}} - \beta_{9,1}c_{NI}^{(m)}, \\
\frac{dc_{NI}}{dt} &= \alpha_{10}c_{NI}^{(m)} - \beta_{10,1}c_{NI} + \gamma_{10,1}c_{NI}(1 - L_I(t)), \\
\frac{dc_T^{(m)}}{dt} &= \alpha_{11}c_L^{g_{11,1}}c_{EC}^{g_{11,2}} - \beta_{11,1}c_T^{(m)}, \\
\frac{dc_T}{dt} &= \alpha_{12}c_T^{(m)} - \beta_{12,1}c_T - \beta_{12,2}c_{ZTL}c_T - \beta_{12,3}c_{ZG}c_T + \gamma_{12,1}c_{ZTL}c_T(1 - L_I(t)) \\
&\quad + \gamma_{12,2}c_{ZG}c_T(1 - L_I(t)), \\
\frac{dc_{E4}^{(m)}}{dt} &= \alpha_{13}c_L^{g_{13,1}}c_{EC}^{g_{13,2}} - \beta_{13,1}c_{E4}^{(m)}, \\
\frac{dc_{E4}}{dt} &= \alpha_{14}c_{E4}^{(m)} - \beta_{14,1}c_{E4} - \beta_{14,2}c_{E4}c_{E3}^{(n)}c_{LUX} - \beta_{14,3}c_{E4}c_{E3}^{(n)}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \alpha_{15}c_L^{g_{15,1}} - \beta_{15,1}c_{E3}^{(m)}, \\
\frac{dc_{E3}^{(c)}}{dt} &= \alpha_{16}c_{E3}^{(m)g_{16,1}} - \beta_{16,1}c_{E3}^{(c)} - \beta_{16,2}c_{E3}^{(c)}c_{COP1}^{(c)} - \beta_{16,3}c_{E3}^{(c)}c_{GI}, \\
\frac{dc_{E3}^{(n)}}{dt} &= \alpha_{17}c_{E3}^{(c)g_{17,1}} - \beta_{17,1}c_{E3}^{(n)} - \beta_{17,2}c_{E3}^{(n)}c_{GI} - \beta_{17,3}c_{E3}^{(n)}c_{COP1}^{(n)} - \beta_{17,4}c_{E3}^{(n)}c_{COP1}^{(n)} \\
&\quad - \beta_{17,5}c_{E3}^{(n)}c_{E4}c_{LUX} - \beta_{17,6}c_{E3}^{(n)}c_{E4}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \alpha_{18}c_L^{g_{18,1}}c_{EC}^{g_{18,2}} - \beta_{18,1}c_{LUX}^{(m)}, \\
\frac{dc_{LUX}}{dt} &= \alpha_{19}c_{LUX}^{(m)} - \beta_{19,1}c_{LUX} - \beta_{19,2}c_{E3}^{(n)}c_{E4}c_{LUX}, \\
\frac{dc_{COP1}^{(c)}}{dt} &= \alpha_{20} - \beta_{20,1}c_{COP1}^{(c)} + \gamma_{20,1}c_{COP1}^{(c)}L_I(t),
\end{aligned}$$

$$\begin{aligned}
\frac{dc_{COP1}^{(n)}}{dt} &= \alpha_{21}c_{COP1}^{(c)g_{21,1}} - \beta_{21,1}c_{COP1}^{(n)} + \gamma_{21,1}c_{COP1}^{(n)}c_PL_I(t) + \gamma_{21,2}c_{COP1}^{(n)}L_I(t), \\
\frac{dc_{COP1d}}{dt} &= \alpha_{22}c_{COP1}^{(n)} - \beta_{22,1}c_{COP1d} + \gamma_{22,1}c_{COP1}^{(n)}c_PL_I(t) + \gamma_{22,2}c_{COP1d}(1 - L_I(t)), \\
\frac{dc_{EG}}{dt} &= \alpha_{23}c_{E3}^{(c)}c_{GI} - \beta_{23,1}c_{EG} - \beta_{23,2}c_{EG}c_{COP1}^{(c)} - \beta_{23,3}c_{GI}c_{E3}^{(n)}, \\
\frac{dc_{EC}}{dt} &= \alpha_{24}c_{E3}^{(n)}c_{E4}c_{LUX} - \beta_{24,1}c_{EC} - \beta_{24,2}c_{EC}c_{COP1}^{(n)} - \beta_{24,3}c_{EC}c_{COP1d} + \gamma_{24,1}c_{EC}c_{GI}L_I(t), \\
\frac{dc_{ZTL}}{dt} &= \alpha_{25} - \beta_{25,1}c_{ZTL} + \gamma_{25,1}c_{ZTL}c_{GI}L_I(t) + \gamma_{25,2}c_{ZG}(1 - L_I(t)), \\
\frac{dc_{ZG}}{dt} &= -\beta_{26,1}c_{ZG} + \gamma_{26,1}c_{GI}c_{ZTL}L_I(t) + \gamma_{26,2}c_{ZG}(1 - L_I(t)), \\
\frac{dc_{GI}^{(m)}}{dt} &= \alpha_{27}c_L^{g_{27,1}}c_{EC}^{g_{27,2}} - \beta_{27,1}c_{GI}^{(m)} + \gamma_{27,1}c_PL_I(t), \\
\frac{dc_{GI}}{dt} &= \alpha_{28}c_{GI}^{(m)} - \beta_{28,1}c_{GI} - \beta_{28,2}c_{GI}c_{E3}^{(c)} + \gamma_{28,1}c_{GI}c_{ZTL}L_I(t) - \gamma_{28,2}c_{ZG}(1 - L_I(t)). \quad (S1.6)
\end{aligned}$$

1.7 KF2014: original equations

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= \frac{LC_{common}}{1 + (r_{11}LC)^2} - m_1c_L^{(m)}, \\
\frac{dc_L}{dt} &= (L_I(t) + m_4(1 - L_I(t)))c_L^{(m)} - m_3c_L, \\
\frac{dc_C^{(m)}}{dt} &= LC_{common} - m_1c_C^{(m)}, \\
\frac{dc_C}{dt} &= (L_I(t) + m_4(1 - L_I(t)))c_C^{(m)} - m_3c_C, \\
\frac{dc_P}{dt} &= p_7(1 - L_I(t))(1 - c_P) - m_{11}c_PL_I(t), \\
\frac{dc_{P9}^{(m)}}{dt} &= q_3c_PL_I(t) - m_{12}c_{P9}^{(m)} \\
&\quad + \frac{1 + a_3r_{33}c_R}{(1 + r_{33}c_R) \left(1 + (r_5LC)^2\right) \left(1 + (r_6EC)^2\right) \left(1 + (r_7c_T^{(n)})^2\right) \left(1 + (r_{40}c_{P5}^{(n)})^2\right)}, \\
\frac{dc_{P9}}{dt} &= c_{P9}^{(m)} - m_{13}c_{P9}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \frac{1}{\left(1 + (r_8LC)^2\right) \left(1 + (r_9EC)^2\right) \left(1 + (r_{10}c_T^{(n)})^2\right) \left(1 + (r_{41}c_{P5}^{(n)})^2\right)} - m_{14}c_{P7}^{(m)}, \\
\frac{dc_{P7}}{dt} &= c_{P7}^{(m)} - (m_{15} + m_{23}(1 - L_I(t)))c_{P7}, \\
\frac{dc_{P5}^{(m)}}{dt} &= \frac{1 + a_4r_{34}c_R}{(1 + r_{34}c_R) \left(1 + (r_{12}LC)^2\right) \left(1 + (r_{13}EC)^2\right) \left(1 + (r_{14}c_T^{(n)})^2\right)} - m_{16}c_{P5}^{(m)}, \\
\frac{dc_{P5}^{(c)}}{dt} &= c_{P5}^{(m)} - (m_{17} + m_{24}c_{ZTL})c_{P5}^{(c)} - P5_{trans}, \\
\frac{dc_{P5}^{(n)}}{dt} &= P5_{trans} - m_{42}c_{P5}^{(n)},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_T^{(m)}}{dt} &= \frac{1 + a_5 r_{35} c_R}{(1 + r_{35} c_R) \left(1 + (r_{15} LC)^2\right) \left(1 + (r_{16} EC)^2\right) \left(1 + \left(r_{17} c_T^{(n)}\right)^2\right)} - m_5 c_T^{(m)}, \\
\frac{dc_T^{(c)}}{dt} &= c_T^{(m)} - (m_8 + m_6 c_{ZTL}) c_T^{(c)} - T_{trans}, \\
\frac{dc_T^{(n)}}{dt} &= T_{trans} - \frac{m_{43}}{1 + m_{38} c_{P5}^{(n)}} c_T^{(n)}, \\
\frac{dc_{E4}^{(m)}}{dt} &= \frac{1 + a_6 r_{36} c_R}{(1 + r_{36} c_R) (1 + (r_{18} EC)^2) (1 + (r_{19} LC)^2) \left(1 + \left(r_{20} c_T^{(n)}\right)^2\right)} - m_7 c_{E4}^{(m)}, \\
\frac{dc_{E4}}{dt} &= p_{23} c_{E4}^{(m)} - m_{35} c_{E4} - c_{E4}^2, \\
\frac{dc_{E4d}}{dt} &= c_{E4}^2 - m_{36} c_{E4d} - E34_{prod}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \frac{1}{1 + (r_{21} LC)^2} - m_{26} c_{E3}^{(m)}, \\
\frac{dc_{E3}}{dt} &= p_{16} c_{E3}^{(m)} - E34_{prod} - E3_{deg} c_{E3}, \\
\frac{dc_{E34}}{dt} &= E34_{prod} - m_{22} c_{E34} E3_{deg}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \frac{1 + a_7 r_{37} c_R}{(1 + r_{37} c_R) \left(1 + (r_{22} EC)^2\right) \left(1 + (r_{23} LC)^2\right) \left(1 + \left(r_{24} c_T^{(n)}\right)^2\right)} - m_{34} c_{LUX}^{(m)}, \\
\frac{dc_{LUX}}{dt} &= c_{LUX}^{(m)} - m_{39} c_{LUX}, \\
\frac{dc_{COP1}^{(c)}}{dt} &= n_5 - p_6 c_{COP1}^{(c)} - m_{27} (1 + p_{15} L_I(t)) c_{COP1}^{(c)}, \\
\frac{dc_{COP1}^{(n)}}{dt} &= p_6 c_{COP1}^{(c)} - (n_{14} + n_6 L_I(t) c_P) c_{COP1}^{(n)} - m_{27} (1 + p_{15} L_I(t)) c_{COP1}^{(n)}, \\
\frac{dc_{COP1d}}{dt} &= (n_{14} + n_6 L_I(t) c_P) c_{COP1}^{(n)} - m_{31} (1 + m_{33} (1 - L_I(t))) c_{COP1d}, \\
\frac{dc_{ZTL}}{dt} &= p_{14} - ZG_{prod} - m_{20} c_{ZTL}, \\
\frac{dc_{ZG}}{dt} &= ZG_{prod} - m_{21} c_{ZG}, \\
\frac{dc_{GI}^{(m)}}{dt} &= \frac{1 + a_8 r_{38} c_R}{(1 + r_{38} c_R) \left(1 + (r_{25} EC)^2\right) \left(1 + (r_{26} LC)^2\right) \left(1 + \left(r_{27} c_T^{(n)}\right)^2\right)} - m_{18} c_G^{(m)}, \\
\frac{dc_{GI}^{(c)}}{dt} &= p_{11} c_G^{(m)} - ZG_{prod} - G_{trans} - m_{19} c_G^{(c)}, \\
\frac{dc_{GI}^{(n)}}{dt} &= G_{trans} - m_{19} c_G^{(n)} - m_{25} c_{E3tot} (1 + m_{28} c_{COP1d} + m_{32} c_{COP1n}) c_G^{(n)}, \\
\frac{dc_{NOX}^{(m)}}{dt} &= \frac{1}{\left(1 + (r_{28} LC)^2\right) \left(1 + (r_{29} c_{P7})^2\right)} - m_{44} c_{NOX}^{(m)},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_{NOX}}{dt} &= c_{NOX}^{(m)} - m_{45}c_{NOX}, \\
\frac{dc_R^{(m)}}{dt} &= \frac{1}{1 + (r_{30}c_{P9})^2 + (r_{31}c_{P7})^2 + \left(r_{32}c_{P5}^{(n)}\right)^2} - m_{46}c_R^{(m)}, \\
\frac{dc_R}{dt} &= c_R^{(m)} - m_{47}c_R, \\
LC &= c_L + f_5c_C, \\
LC_{common} &= \frac{q_1L_I(t)c_P + 1}{1 + (r_{1c_{P9}})^2 + (r_{2c_{P7}})^2 + \left(r_{3c_{P5}}^{(n)}\right)^2 + \left(r_{4c_T}^{(n)}\right)^2}, \\
EC &= \frac{(c_{LUX} + f_6c_{NOX})(c_{E34} + f_1c_{E3})}{1 + f_3(c_{LUX} + f_2c_{NOX}) + f_4(c_{E34} + f_1c_{E3})}, \\
P5_{trans} &= t_5c_{P5}^{(c)} - t_6c_{P5}^{(n)}, \\
T_{trans} &= t_7c_T^{(c)} - \frac{t_8}{1 + m_{37}c_{P5}^{(n)}}c_T^{(n)}, \\
E34_{prod} &= p_{25}c_{E3}c_{E4d}, \\
E3_{deg} &= m_{30}c_{COP1d} + m_{29}c_{COP1n} + m_9 + m_{10}c_G^{(n)}, \\
ZG_{prod} &= p_{12}c_{ZTL}c_G^{(c)} - (p_{13}(1 - L_I(t)) + p_{10}L_I(t))c_{ZG}, \\
c_{E3tot} &= c_{E3} + c_{E34}, \\
G_{trans} &= p_{28}c_G^{(c)} - \frac{p_{29}}{1 + t_9c_{E3tot}}c_G^{(n)}. \tag{S1.7}
\end{aligned}$$

1.8 KF2014S: extended S-System formulation

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= \alpha_1 \left(c_{P9} + c_{P7} + c_{P5}^{(n)} + c_T^{(n)} \right)^{g_{1,1}} - \beta_{1,1}c_L^{(m)} + \gamma_{1,1}c_PL_I(t), \\
\frac{dc_L}{dt} &= \alpha_2c_L^{(m)} - \beta_{2,1}c_L + \gamma_{2,1}c_L^{(m)}L_I(t), \\
\frac{dc_C^{(m)}}{dt} &= \alpha_3 \left(c_{P9} + c_{P7} + c_{P5}^{(n)} + c_T^{(n)} \right)^{g_{3,1}} - \beta_{3,1}c_C^{(m)} + \gamma_{3,1}c_PL_I(t), \\
\frac{dc_C}{dt} &= \alpha_4c_C^{(m)} - \beta_{4,1}c_C + \gamma_{4,1}c_C^{(m)}L_I(t), \\
\frac{dc_P}{dt} &= \gamma_{5,1}(1 - L_I(t)) + \gamma_{5,2}c_P(1 - L_I(t)) + \gamma_{5,3}c_PL_I(t), \\
\frac{dc_{P9}^{(m)}}{dt} &= \alpha_6c_{R8}^{g_{6,1}}(c_L + c_C)^{g_{6,2}}(c_{LUX} + c_{NOX})^{g_{6,3}}(c_{E3} + c_{E34})^{g_{6,4}}c_T^{(n)g_{6,5}}c_{P5}^{(n)g_{6,6}} \\
&\quad - \beta_{6,1}c_{P9}^{(m)} + \gamma_{6,1}c_PL_I(t), \\
\frac{dc_{P9}}{dt} &= \alpha_7c_{P9}^{(m)} - \beta_{7,1}c_{P9}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \alpha_8(c_L + c_C)^{g_{8,1}}(c_{LUX} + c_{NOX})^{g_{8,2}}(c_{E3} + c_{E34})^{g_{8,3}}c_T^{(n)g_{8,4}}c_{P5}^{(n)g_{8,5}} - \beta_{8,1}c_{P7}^{(m)}, \\
\frac{dc_{P7}}{dt} &= \alpha_9c_{P7}^{(m)} - \beta_{9,1}c_{P7} + \gamma_{9,1}c_{P7}(1 - L_I(t)), \\
\frac{dc_{P5}^{(m)}}{dt} &= \alpha_{10}c_{R8}^{g_{10,1}}(c_L + c_C)^{g_{10,2}}(c_{LUX} + c_{NOX})^{g_{10,3}}(c_{E3} + c_{E34})^{g_{10,4}}c_T^{(n)g_{10,5}} - \beta_{10,1}c_{P5}^{(m)},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_{P5}^{(c)}}{dt} &= \alpha_{11}c_{P5}^{(m)g_{11,1}} - \beta_{11,1}c_{P5}^{(c)} - \beta_{11,2}c_{ZTL}c_{P5}^{(c)}, \\
\frac{dc_{P5}^{(n)}}{dt} &= \alpha_{12}c_{P5}^{(c)g_{12,1}} - \beta_{12,1}c_{P5}^{(n)}, \\
\frac{dc_T^{(m)}}{dt} &= \alpha_{13}c_{R8}^{g_{13,1}}(c_L + c_C)^{g_{13,2}}(c_{LUX} + c_{NOX})^{g_{13,3}}(c_{E3} + c_{E34})^{g_{13,4}}c_T^{(n)g_{13,5}} - \beta_{13,1}c_T^{(m)}, \\
\frac{dc_T^{(c)}}{dt} &= \alpha_{14}c_T^{(m)g_{14,1}} - \beta_{14,1}c_T^{(c)} - \beta_{14,2}c_{ZTL}c_T^{(c)}, \\
\frac{dc_T^{(n)}}{dt} &= \alpha_{15}c_T^{(c)g_{15,1}} - \beta_{15,1}c_T^{(n)}, \\
\frac{dc_{E4}^{(m)}}{dt} &= \alpha_{16}c_{R8}^{g_{16,1}}(c_L + c_C)^{g_{16,2}}(c_{LUX} + c_{NOX})^{g_{16,3}}(c_{E3} + c_{E34})^{g_{16,4}}c_T^{(n)g_{16,5}} - \beta_{16,1}c_{E4}^{(m)}, \\
\frac{dc_{E4}}{dt} &= \alpha_{17}c_{E4}^{(m)} - \beta_{17,1}c_{E4}, \\
\frac{dc_{E4d}}{dt} &= \alpha_{18}c_{E4} - \beta_{18,1}c_{E4d} - \beta_{18,2}c_{E3}c_{E4d}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \alpha_{19}(c_L + c_C)^{g_{19,1}} - \beta_{19,2}c_{E3}, \\
\frac{dc_{E3}}{dt} &= \alpha_{20}c_{E3}^{(m)} - \beta_{20,1}c_{E3} - \beta_{20,2}c_{E3}c_{E4d} - \beta_{20,3}c_{E3}c_{COP1}^{(n)} - \beta_{20,4}c_{E3}c_{COP1d} - \beta_{20,5}c_{E3}c_{GI}^{(n)}, \\
\frac{dc_{E34}}{dt} &= \alpha_{21}c_{E3}c_{E4d} - \beta_{21,1}c_{E34} - \beta_{21,2}c_{E34}c_{COP1d} - \beta_{21,3}c_{E34}c_{COP1}^{(n)} - \beta_{21,4}c_{E34}c_{GI}^{(n)}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \alpha_{22}c_{R8}^{g_{22,1}}(c_L + c_C)^{g_{22,2}}(c_{LUX} + c_{NOX})^{g_{22,3}}(c_{E3} + c_{E34})^{g_{22,4}}c_T^{(n)g_{22,5}} - \beta_{22,1}c_{LUX}^{(m)}, \\
\frac{dc_{LUX}}{dt} &= \alpha_{23}c_{LUX}^{(m)} - \beta_{23,1}c_{LUX}, \\
\frac{dc_{COP1}^{(c)}}{dt} &= \alpha_{24} - \beta_{24,1}c_{COP1}^{(c)} + \gamma_{24,1}c_{COP1}^{(c)}L_I(t), \\
\frac{dc_{COP1}^{(n)}}{dt} &= \alpha_{25}c_{COP1}^{(c)g_{25,1}} - \beta_{25,1}c_{COP1}^{(n)} + \gamma_{25,1}c_{PC}c_{COP1}^{(n)}L_I(t) + \gamma_{25,2}c_{COP1}^{(n)}L_I(t), \\
\frac{dc_{COP1d}}{dt} &= \alpha_{26}c_{COP1}^{(n)} - \beta_{26,1}c_{COP1d} + \gamma_{26,1}c_{PC}c_{COP1}^{(n)}L_I(t) + \gamma_{26,2}c_{COP1d}(1 - L_I(t)), \\
\frac{dc_{ZTL}}{dt} &= \alpha_{27} - \beta_{27,1}c_{ZTL} - \beta_{27,2}c_{ZTL}c_{GI}^{(c)} + \gamma_{27,1}c_{ZG}L_I(t) + \gamma_{27,2}c_{ZG}(1 - L_I(t)), \\
\frac{dc_{ZG}}{dt} &= \alpha_{28}c_{ZTL}c_{GI}^{(c)} - \beta_{28,1}c_{ZG} + \gamma_{28,1}c_{ZG}L_I(t) + \gamma_{28,2}c_{ZG}(1 - L_I(t)), \\
\frac{dc_{GI}^{(m)}}{dt} &= \alpha_{29}c_{R8}^{g_{29,1}}(c_L + c_C)^{g_{29,2}}(c_{LUX} + c_{NOX})^{g_{29,3}}(c_{E3} + c_{E34})^{g_{29,4}}c_T^{(n)g_{29,5}} - \beta_{29,1}c_{GI}^{(m)}, \\
\frac{dc_{GI}^{(c)}}{dt} &= \alpha_{30}c_{GI}^{(m)g_{30,1}} - \beta_{30,1}c_{GI}^{(c)} - \beta_{30,2}c_{ZTL}c_{GI}^{(c)} + \gamma_{30,1}c_{ZG}L_I(t) + \gamma_{30,2}c_{ZG}(1 - L_I(t)), \\
\frac{dc_{GI}^{(n)}}{dt} &= \alpha_{31}c_{GI}^{(c)g_{31,1}} - \beta_{31,1}c_{GI}^{(n)} - \beta_{31,2}(c_{E3} + c_{E34})c_{GI}^{(n)} - \beta_{31,3}(c_{E3} + c_{E34})c_{COP1}^{(n)}c_{GI}^{(n)} \\
&\quad - \beta_{31,4}(c_{E3} + c_{E34})c_{COP1d}c_{GI}^{(n)}, \\
\frac{dc_{NOX}^{(m)}}{dt} &= \alpha_{32}(c_L + c_C)^{g_{32,1}}c_{P7}^{g_{32,2}} - \beta_{32,1}c_{NOX}^{(m)}, \\
\frac{dc_{NOX}}{dt} &= \alpha_{33}c_{NOX}^{(m)} - \beta_{33,1}c_{NOX},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_{R8}^{(m)}}{dt} &= \alpha_{34} \left(c_{P9} + c_{P7} + c_{P5}^{(n)} \right)^{g_{34,1}} - \beta_{34,1} c_{R8}^{(m)}, \\
\frac{dc_{R8}}{dt} &= \alpha_{35} c_{R8}^{(m)} - \beta_{35,1} c_{R8}.
\end{aligned} \tag{S1.8}$$

1.9 MF2016K: original equations

$$\begin{aligned}
\frac{dc_P^{(p)}}{dt} &= \phi_1 L_I(t) + \phi_2 c_P^{(p)} + \phi_3, \\
\frac{dc_L^{(m)}}{dt} &= \frac{\phi_4}{\phi_5^2 + \phi_6 \left(\phi_7 c_{P9}^{(p)} + \phi_8 c_{P7}^{(p)} + \phi_9 c_{P5}^{(p)} \right)^2} + \phi_{10} c_L^{(m)}, \\
\frac{dc_L^{(p)}}{dt} &= \phi_{11} c_L^{(m)} + \phi_{12} c_L^{(p)}, \\
\frac{dc_{P9}^{(m)}}{dt} &= \phi_{13} c_P^{(p)} L_I(t) + \phi_{14} + \frac{\phi_{15} c_L^{(p)2}}{\phi_{16}^2 + c_L^{(p)2}} + \phi_{17} c_{P9}^{(m)}, \\
\frac{dc_{P9}^{(p)}}{dt} &= \phi_{18} c_{P9}^{(m)} + \phi_{19} c_{P9}^{(p)}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \frac{\phi_{20}}{\left(\phi_{21} + c_{EC}^{(p)} \right) \left(\phi_{22}^2 + c_T^{(p)2} \right)} + \phi_{23} c_{P7}^{(m)}, \\
\frac{dc_{P7}^{(p)}}{dt} &= \phi_{24} c_{P7}^{(m)} + \phi_{25} c_{P7}^{(p)}, \\
\frac{dc_{P5}^{(m)}}{dt} &= \frac{\phi_{26}}{\left(\phi_{27}^2 + c_T^{(p)2} \right) \left(\phi_{28}^2 + c_L^{(p)2} \right)} + \phi_{29} c_{P5}^{(m)}, \\
\frac{dc_{P5}^{(p)}}{dt} &= \phi_{30} c_{P5}^{(m)} + \phi_{31} c_{P5}^{(p)}, \\
\frac{dc_T^{(m)}}{dt} &= \frac{\phi_{32}}{\left(\phi_{33} + c_{EC}^{(p)} \right) \left(\phi_{34}^2 + c_L^{(p)2} \right)} + \phi_{35} c_T^{(m)}, \\
\frac{dc_T^{(p)}}{dt} &= \phi_{36} c_T^{(m)} + \phi_{37} c_T^{(p)}, \\
\frac{dc_{EC}^{(p)}}{dt} &= \phi_{38} c_{E3}^{(p)} c_{E4}^{(p)} c_{LUX}^{(p)} + \phi_{39} c_{EC}^{(p)}, \\
\frac{dc_{R8}^{(m)}}{dt} &= \frac{\phi_{40}}{\phi_{41}^2 + \phi_9 c_{P5}^{(p)2}} + \phi_{42} c_{R8}^{(m)}, \\
\frac{dc_{R8}^{(p)}}{dt} &= \phi_{43} c_{R8}^{(m)} + \phi_{44} c_{R8}^{(p)}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \frac{\phi_{45}}{\phi_{46}^2 + c_L^{(p)2}} + \phi_{47} c_{E3}^{(m)}, \\
\frac{dc_{E3}^{(p)}}{dt} &= \phi_{48} c_{E3}^{(m)} + \phi_{49} c_{E3}^{(p)},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_{E4}^{(m)}}{dt} &= \frac{\phi_{50}}{\left(\phi_{51} + c_{EC}^{(p)}\right) \left(\phi_{52}^2 + c_L^{(p)2}\right)} + \phi_{53} c_{E4}^{(m)}, \\
\frac{dc_{E4}^{(p)}}{dt} &= \phi_{54} c_{E4}^{(m)} + \phi_{55} c_{E4}^{(p)}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \frac{\phi_{56}}{\left(\phi_{57} + c_{EC}^{(p)}\right) \left(\phi_{58}^2 + c_L^{(p)2}\right)} + \phi_{59} c_{LUX}^{(m)}, \\
\frac{dc_{LUX}^{(p)}}{dt} &= \phi_{60} c_{LUX}^{(m)} + \phi_{61} c_{LUX}^{(p)}, \\
\frac{dc_{GI}^{(m)}}{dt} &= \frac{\phi_{62}}{\left(\phi_{63}^2 + c_L^{(p)2}\right) \left(\phi_{64}^2 + c_T^{(p)2}\right) \left(\phi_{65}^2 + c_{E3}^{(p)2}\right)} + \phi_{66} c_{GI}^{(m)}, \\
\frac{dc_{GI}^{(p)}}{dt} &= \phi_{67} c_{GI}^{(m)} + \phi_{68} c_{COP1}^{(p)} c_{E3}^{(p)} c_{GI}^{(p)} + \phi_{69} c_{GI}^{(p)}, \\
\frac{dc_{COP1}^{(p)}}{dt} &= \phi_{70} L_I(t) + \phi_{71} c_{COP1}^{(p)} + \phi_{72}, \\
\frac{dc_{ZTL}^{(p)}}{dt} &= \frac{\phi_{73} c_{ZTL}^{(p)} c_{GI}^{(p)}}{1 + \theta_{144} c_{ZTL}^{(p)} + \theta_{145} c_{GI}^{(p)}} + \phi_{74} c_{ZTL}^{(p)} + \phi_{75}.
\end{aligned} \tag{S1.9}$$

1.10 MF2016KS: extended S-System formulation

$$\begin{aligned}
\frac{dc_P}{dt} &= \alpha_1 - \beta_{1,1} c_P + \gamma_{1,1} L_I(t), \\
\frac{dc_L^{(m)}}{dt} &= \alpha_2 (c_{P9} + c_{P7} + c_{P5})^{g_{2,1}} - \beta_{2,1} c_L^{(m)}, \\
\frac{dc_L}{dt} &= \alpha_3 c_L^{(m)} - \beta_{3,1} c_L, \\
\frac{dc_{P9}^{(m)}}{dt} &= \alpha_4 c_L^{g_{4,1}} - \beta_{4,1} c_{P9}^{(m)} + \gamma_{4,1} c_P L_I(t), \\
\frac{dc_{P9}}{dt} &= \alpha_5 c_{P9}^{(m)} - \beta_{5,1} c_{P9}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \alpha_6 c_T^{g_{6,1}} c_{EC}^{g_{6,2}} - \beta_{6,1} c_{P7}^{(m)}, \\
\frac{dc_{P7}}{dt} &= \alpha_7 c_{P7}^{(m)} - \beta_{7,1} c_{P7}, \\
\frac{dc_{P5}^{(m)}}{dt} &= \alpha_8 c_L^{g_{8,1}} c_T^{g_{8,2}} - \beta_{8,1} c_{P5}^{(m)}, \\
\frac{dc_{P5}}{dt} &= \alpha_9 c_{P5}^{(m)} - \beta_{9,1} c_{P5}, \\
\frac{dc_T^{(m)}}{dt} &= \alpha_{10} c_L^{g_{10,1}} c_{EC}^{g_{10,2}} - \beta_{10,1} c_T^{(m)}, \\
\frac{dc_T}{dt} &= \alpha_{11} c_T^{(m)} - \beta_{11,1} c_T, \\
\frac{dc_{EC}}{dt} &= \alpha_{12} c_{E3} c_{E4} c_{LUX} - \beta_{12,1} c_{EC}, \\
\frac{dc_{R8}^{(m)}}{dt} &= \alpha_{13} c_{P5}^{g_{13,1}} - \beta_{13,1} c_{R8},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_{R8}}{dt} &= \alpha_{14}c_{R8}^{(m)} - \beta_{14,1}c_{R8}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \alpha_{15}c_L^{g_{15,1}} - \beta_{15,1}c_{E3}^{(m)}, \\
\frac{dc_{E3}}{dt} &= \alpha_{16}c_{E3}^{(m)} - \beta_{16,1}c_{E3}, \\
\frac{dc_{E4}^{(m)}}{dt} &= \alpha_{17}c_L^{g_{17,1}}c_{EC}^{g_{17,2}} - \beta_{17,1}c_{E4}^{(m)}, \\
\frac{dc_{E4}}{dt} &= \alpha_{18}c_{E4}^{(m)} - \beta_{18,1}c_{E4}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \alpha_{19}c_L^{g_{19,1}}c_{EC}^{g_{19,2}} - \beta_{19,1}c_{LUX}^{(m)}, \\
\frac{dc_{LUX}}{dt} &= \alpha_{20}c_{LUX}^{(m)} - \beta_{20,1}c_{LUX}, \\
\frac{dc_{GI}^{(m)}}{dt} &= \alpha_{21}c_L^{g_{21,1}}c_T^{g_{21,2}}c_{E3}^{g_{21,3}} - \beta_{21,1}c_{GI}^{(m)}, \\
\frac{dc_{GI}}{dt} &= \alpha_{22}c_{GI}^{(m)} - \beta_{22,1}c_{GI} - \beta_{22,2}c_{GI}c_{E3}c_{COP1}, \\
\frac{dc_{COP1}}{dt} &= \alpha_{23} - \beta_{23,1}c_{COP1} + \gamma_{23,1}L_I(t), \\
\frac{dc_{ZTL}}{dt} &= \alpha_{24} - \beta_{24,1}c_{ZTL} - \beta_{24,2}c_{ZTL}^{h_{24,1}}c_{GI}^{h_{24,2}}. \tag{S1.10}
\end{aligned}$$

Note: For the ZTL protein equation, we have included exponents $h_{i,k}$ controlling the stabilisation of ZTL by GI protein. This is because in MF2016K, this process was modelled as a rational function of the two protein concentrations (*cf.* eqs. (S1.9)), rather than as a product of the concentrations, as is usually the case (*cf.* eq. (13)).

1.11 MF2016Korig: original S-System formulation

$$\begin{aligned}
\frac{dc_P}{dt} &= \alpha_1 - \beta_{1,1}L_I(t)^{h_{1,1}}c_P^{h_{1,2}}, \\
\frac{dc_L^{(m)}}{dt} &= \alpha_2c_{P9}^{g_{2,1}}c_{P7}^{g_{2,2}}c_{P5}^{g_{2,3}} - \beta_{2,1}c_L^{(m)}, \\
\frac{dc_L}{dt} &= \alpha_3c_L^{(m)} - \beta_{3,1}c_L, \\
\frac{dc_{P9}^{(m)}}{dt} &= \alpha_4c_L^{g_{4,1}}c_P^{g_{4,2}}L_I(t)^{g_{4,3}} - \beta_{4,1}c_{P9}^{(m)}, \\
\frac{dc_{P9}}{dt} &= \alpha_5c_{P9}^{(m)} - \beta_{5,1}c_{P9}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \alpha_6c_T^{g_{6,1}}c_{EC}^{g_{6,2}} - \beta_{6,1}c_{P7}^{(m)}, \\
\frac{dc_{P7}}{dt} &= \alpha_7c_{P7}^{(m)} - \beta_{7,1}c_{P7}, \\
\frac{dc_{P5}^{(m)}}{dt} &= \alpha_8c_L^{g_{8,1}}c_T^{g_{8,2}} - \beta_{8,1}c_{P5}^{(m)}, \\
\frac{dc_{P5}}{dt} &= \alpha_9c_{P5}^{(m)} - \beta_{9,1}c_{P5}, \\
\frac{dc_T^{(m)}}{dt} &= \alpha_{10}c_L^{g_{10,1}}c_{EC}^{g_{10,2}} - \beta_{10,1}c_T^{(m)},
\end{aligned}$$

$$\begin{aligned}
\frac{dc_T}{dt} &= \alpha_{11}c_T^{(m)} - \beta_{11,1}c_T, \\
\frac{dc_{EC}}{dt} &= \alpha_{12}c_{E3}c_{E4}c_{LUX} - \beta_{12,1}c_{EC}, \\
\frac{dc_{R8}^{(m)}}{dt} &= \alpha_{13}c_{P5}^{g_{13,1}} - \beta_{13,1}c_{R8}, \\
\frac{dc_{R8}}{dt} &= \alpha_{14}c_{R8}^{(m)} - \beta_{14,1}c_{R8}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \alpha_{15}c_L^{g_{15,1}} - \beta_{15,1}c_{E3}^{(m)}, \\
\frac{dc_{E3}}{dt} &= \alpha_{16}c_{E3}^{(m)} - \beta_{16,1}c_{E3}, \\
\frac{dc_{E4}^{(m)}}{dt} &= \alpha_{17}c_L^{g_{17,1}}c_{EC}^{g_{17,2}} - \beta_{17,1}c_{E4}^{(m)}, \\
\frac{dc_{E4}}{dt} &= \alpha_{18}c_{E4}^{(m)} - \beta_{18,1}c_{E4}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \alpha_{19}c_L^{g_{19,1}}c_{EC}^{g_{19,2}} - \beta_{19,1}c_{LUX}^{(m)}, \\
\frac{dc_{LUX}}{dt} &= \alpha_{20}c_{LUX}^{(m)} - \beta_{20,1}c_{LUX}, \\
\frac{dc_{GI}^{(m)}}{dt} &= \alpha_{21}c_L^{g_{21,1}}c_T^{g_{21,2}}c_{E3}^{g_{21,3}} - \beta_{21,1}c_{GI}^{(m)}, \\
\frac{dc_{GI}}{dt} &= \alpha_{22}c_{GI}^{(m)} - \beta_{22,1}c_{GI}^{h_{22,1}}c_{E3}^{h_{22,2}}c_{COP1}^{h_{22,3}}, \\
\frac{dc_{COP1}}{dt} &= \alpha_{23} - \beta_{23,1}L_I(t)^{h_{23,1}}c_{COP1}^{h_{23,2}}, \\
\frac{dc_{ZTL}}{dt} &= \alpha_{24}c_{ZTL}^{g_{24,1}}c_{GI}^{g_{24,2}} - \beta_{24,1}c_{ZTL}.
\end{aligned} \tag{S1.11}$$

2 The sine-sweeping method

In sine-sweeping, the input to the system is a sinusoidal signal whose frequency is varied within the range of interest. By collecting the magnitude and phase values of the output responses, the transfer function of the system can be approximated. Here, a summary of the method is given. For more details see [S2].

Let the sinusoidal input signal be $x(t) = U_{in} \sin(\omega t)$, where U_{in} and ω are the amplitude and frequency respectively. According to linear systems theory, if the system is linear time invariant, the output response will be a sinusoid of the same frequency, but with scaled amplitude and a phase shift. In practice, nonlinearities, transients, and disturbances $V(t)$ will also affect the output. Thus, we can write the output $y(t)$ as

$$y(t) = U_{out} \sin(\omega t + \Phi) + V(t) + \text{nonlinearities} + \text{transients}, \tag{S2.12}$$

where $U_{out} = U_{in}|G(i\omega)|$, $\Phi = \angle G(i\omega) = \arctan\left(\frac{\Im[G(i\omega)]}{\Re[G(i\omega)]}\right)$ and $G(i\omega)$ is the transfer function relating the input and output.

By neglecting the initial part of the data and assuming that the linear dynamics dominate, we can ignore the effects of transients and nonlinearities, respectively. Furthermore, the effect of disturbances $V(t)$ can be reduced through the correlation method [S2], where the idea is to correlate the output response $y(t)$ with a sine and cosine of the same frequency, and then average over the signal length (S9 Fig). Writing the sampled output signal as $\{y(t_k) : 1 \leq k \leq L\}$, the resulting

averages are given below:

$$Q_S(L) = \frac{1}{L} \sum_{k=1}^L y(t_k) \sin(\omega t_k), \quad Q_C(L) = \frac{1}{L} \sum_{k=1}^L y(t_k) \cos(\omega t_k). \quad (\text{S2.13})$$

By substituting (S2.12) into (S2.13) and performing some algebraic manipulation, we obtain

$$\begin{aligned} Q_S(L) &= \frac{U_{in}}{2} |G(i\omega)| \cos \Phi - \frac{U_i}{2} |G(i\omega)| \frac{1}{L} \sum_{k=1}^L \cos(2\omega t_k + \Phi) + \frac{1}{L} \sum_{k=1}^L V(t_k) \sin(\omega t_k), \\ Q_C(L) &= \frac{U_{in}}{2} |G(i\omega)| \sin \Phi - \frac{U_{in}}{2} |G(j\omega)| \frac{1}{L} \sum_{k=1}^L \sin(2\omega t_k + \Phi) + \frac{1}{L} \sum_{k=1}^L V(t_k) \cos(\omega t_k). \end{aligned} \quad (\text{S2.14})$$

The second terms in the expressions for $Q_S(L)$ and $Q_C(L)$ converge to zero as $L \rightarrow \infty$. Furthermore, if $V(t)$ is a zero-mean stationary stochastic process with covariance function $R_c(\tau)$ such that $\sum_{\tau=0}^{\infty} \tau |R_c(\tau)| < \infty$, then it can be shown that the third terms in the expressions for $Q_S(L)$ and $Q_C(L)$ also converge to zero as $L \rightarrow \infty$ (see [S2] for complete details). From the remaining terms in (S2.14), the magnitude, $|G(i\omega)|$, and phase, $\angle G(i\omega)$, can therefore be estimated using the equations below:

$$|G(i\omega)| = \frac{2\sqrt{Q_S(L)^2 + Q_C(L)^2}}{U_{in}}, \quad \angle G(i\omega) = \arctan\left(\frac{Q_C(L)}{Q_S(L)}\right). \quad (\text{S2.15})$$

Plotting $|G(i\omega)|$ and $\angle G(i\omega)$ against frequency ω then yields the Bode plot representation of the sine-sweeping approximation to the transfer function $G(i\omega)$.

3 Transfer function analysis of protein shuttling in JL2006

The Bode plots for JL2006 generated using sine-sweeping that are shown in Fig 6 of the main paper imply that the linear systems approximations to the functions relating input mRNAs to output nuclear proteins can be accurately represented as first-order transfer functions. In this section, we provide further evidence for the viability of this approximation using frequency response analysis, and also show that a first-order transfer function is equivalent to having an aggregated protein equation, rather than two separate equations modelling protein shuttling.

We start by noting that the protein shuttling mechanisms used in JL2006 all have the general form shown below (cf. eqs. (S1.3)):

$$\begin{aligned} \frac{dP_{j,cy}}{dt} &= \alpha G_i - \beta_{cy} P_{j,cy} + \beta_{nu} P_{j,nu} - \gamma_{cy} \frac{P_{j,cy}}{K_{cy} + P_{j,cy}}, \\ \frac{dP_{j,nu}}{dt} &= \beta_{cy} P_{j,cy} - \beta_{nu} P_{j,nu} - \gamma_{nu} \frac{P_{j,nu}}{K_{nu} + P_{j,cy}}. \end{aligned} \quad (\text{S3.1})$$

Here, G_i is the gene, $P_{j,cy}$ is cytoplasmic protein, $P_{j,nu}$ is nuclear protein, α is the translation rate, $\{\gamma_{cy}, \gamma_{nu}\}$ are degradation rates, $\{K_{cy}, K_{nu}\}$ are Michaelis constants for degradation and $\{\beta_{cy}, \beta_{nu}\}$ are transport rates. Linearising the nonlinear degradation terms in eqs. (S3.1) and rescaling the degradation rates yields the following:

$$\begin{aligned} \frac{dP_{j,cy}}{dt} &= \alpha G_i - \beta_{cy} P_{j,cy} + \beta_{nu} P_{j,nu} - \gamma_{cy} P_{j,cy}, \\ \frac{dP_{j,nu}}{dt} &= \beta_{cy} P_{j,cy} - \beta_{nu} P_{j,nu} - \gamma_{nu} P_{j,nu}. \end{aligned} \quad (\text{S3.2})$$

Taking Laplace transforms of (S3.1) and assuming zero initial conditions, we arrive at the following equations involving the transforms of $G_i(t)$, $P_{j,cy}(t)$ and $P_{j,nu}(t)$:

$$\begin{aligned} P_{j,cy}(s) &= \frac{\alpha}{s + \beta_{cy} + \gamma_{cy}} G_i(s) + \frac{\beta_{nu}}{s + \beta_{cy} + \gamma_{cy}} P_{j,nu}(s), \\ P_{j,nu}(s) &= \frac{\beta_{cy}}{s + \beta_{nu} + \gamma_{nu}} P_{j,cy}(s). \end{aligned} \quad (\text{S3.3})$$

Some algebraic manipulation then shows that the transfer function relating input mRNA ($G_i(s)$) to output nuclear protein ($P_{j,nu}(s)$) is given by:

$$\frac{P_{j,nu}(s)}{G_i(s)} = \frac{\alpha\beta_{cy}}{s^2 + (\beta_{cy} + \gamma_{cy} + \beta_{nu} + \gamma_{nu})s + (\beta_{nu}\gamma_{cy} + \beta_{cy}\gamma_{nu} + \gamma_{cy}\gamma_{nu})}. \quad (\text{S3.4})$$

We note that eq. (S3.4) is a second order-transfer function. The only way in which eq. (S3.4) can be approximated by a first-order transfer function, as suggested by the sine-sweeping results, is if its two poles (the roots of its denominator) are far apart; in other words, if eq. (S3.4) has one fast pole and one slow pole.

We demonstrate that is indeed the case for component Y of JL2006, and show the similarity between the Bode plots for the full second-order transfer function and its first-order approximation (similar results can be demonstrated for the other components of the model). In order to compute the full second-order transfer function of Y, we take the values $\alpha = p_4 = 0.2485$, $\beta_{cy} = r_7 = 2.2123$ and $\beta_{nu} = r_8 = 0.2002$ from **S5 Table**. For the nonlinear degradation terms, we use the best linear approximations (see **S10 Fig**), yielding the values $\gamma_{cy} = 0.073$ and $\gamma_{nu} = 0.324$. Substituting all these values into eq. (S3.4) gives

$$\frac{P_{j,nu}(s)}{G_i(s)} = \frac{0.04975}{s^2 + 2.809s + 0.25} = \frac{0.04975}{(s + 0.092)(s + 2.718)}. \quad (\text{S3.5})$$

Note that the two poles differ in value by almost a factor of 30; there is thus one fast pole and one slow pole, as predicted. The Bode plot for eq. (S3.5), plotted over the same frequency range as that used for sine-sweeping, is shown as blue lines in **S11 Fig**.

We now consider the equation for an aggregated model of nuclear protein production (*cf.* eq. (10) of the main paper):

$$\frac{dP_{j,nu}}{dt} = aG_i - bP_{j,nu}. \quad (\text{S3.6})$$

Here, a and b represent translation and degradation, respectively. Taking Laplace transforms and assuming zero initial conditions as before results in the following first-order transfer function:

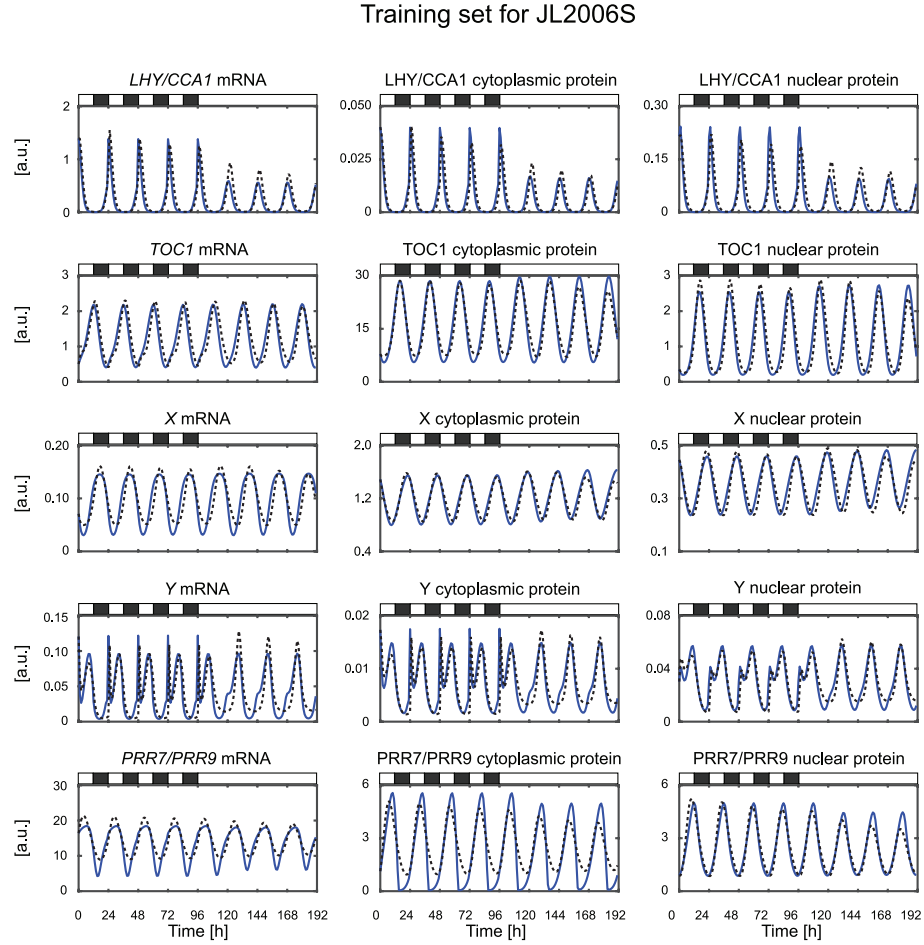
$$\frac{P_{j,nu}(s)}{G_i(s)} = \frac{a}{s + b}. \quad (\text{S3.7})$$

The values of a and b in eq. (S3.7) can be inferred from a Bode plot by using the fact that b is the corner frequency – the frequency at which the magnitude trace starts to deviate from its plateau value $A_{vdB} = 20 \log_{10}(a/b)$. The Bode plot for eq. (S3.5) implies $b \approx 0.090$ and $A_{vdB} = -14$ dB, from which it follows that $a \approx 0.018$. Comparing Bode plots in **S11 Fig** shows that the full second-order transfer function for Y (eq. (S3.5)) is quite accurately modelled by the first-order approximation (eq. (S3.7)) with these values of a and b , particularly in the lower frequency range.

References

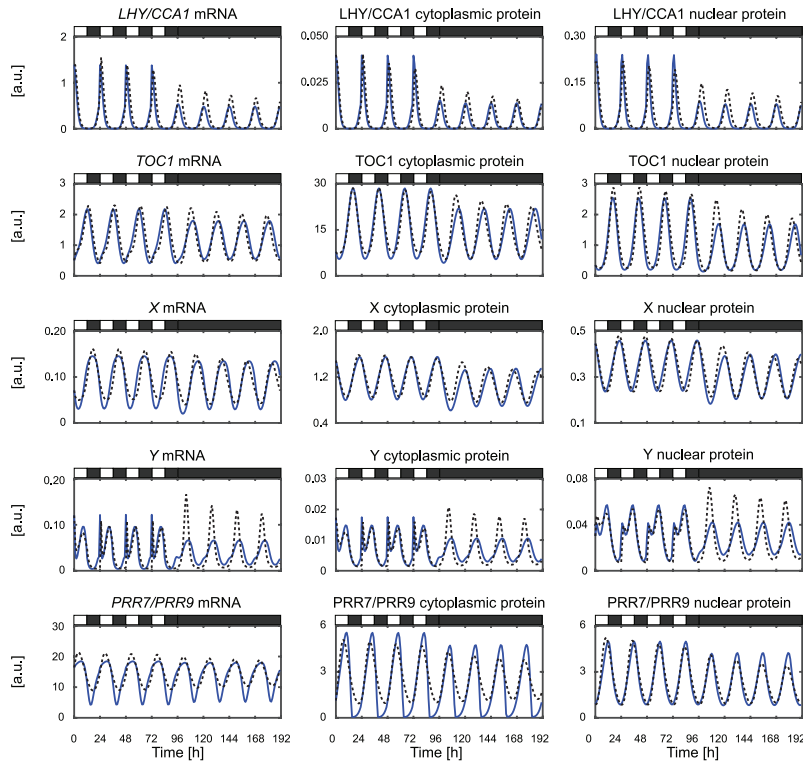
- [S1] Locke JCW, Millar AJ, Turner MS. Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. J Theor Biol. 2005;234:383–393. doi:10.1016/j.jtbi.2004.11.038.
- [S2] Ljung L. System Identification: Theory for the User. 2nd ed. Prentice Hall; 1999.

Supplementary Figures

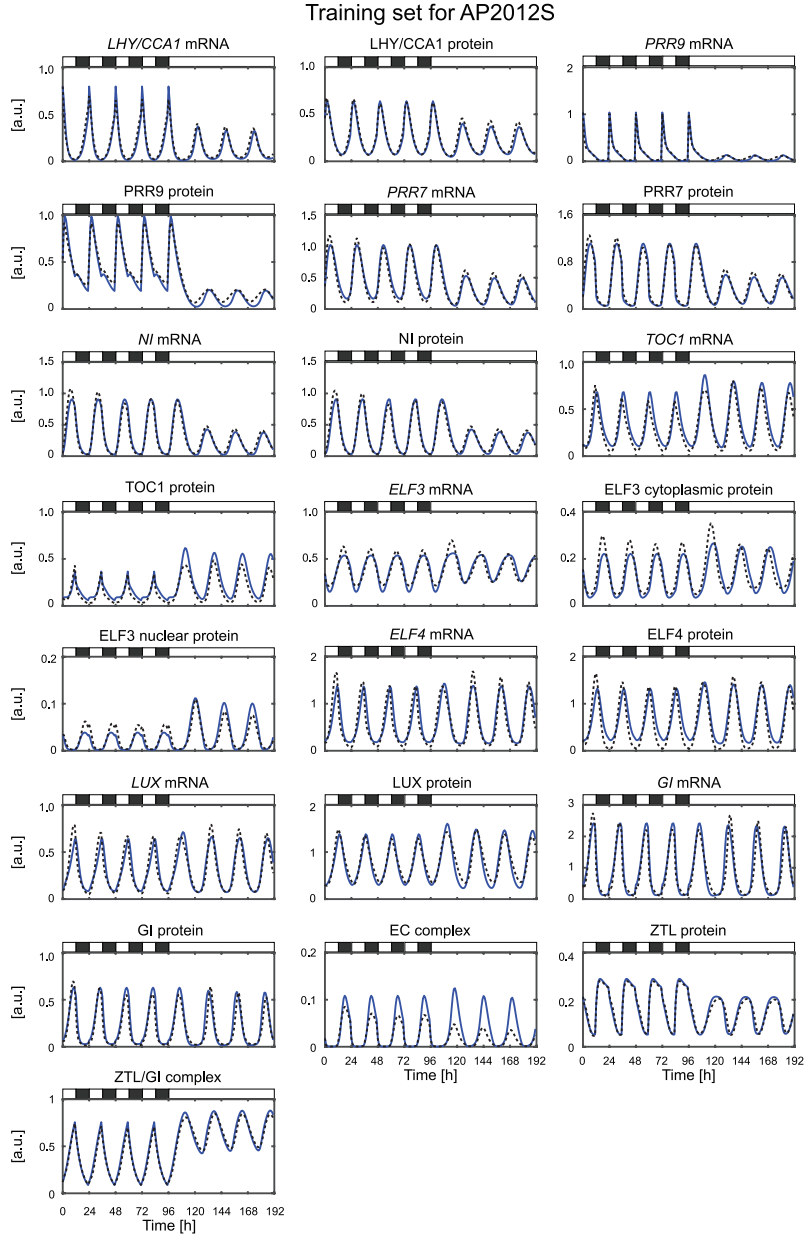


S1 Fig. Plant clock model JL2006S – optimal fits to synthetic training data. Blue solid lines show timeseries generated by JL2006 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant light (LL). Black dashed lines show timeseries obtained by optimising the parameters of JL2006S to this data in the same simulated light environment (see **Fig 2B** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

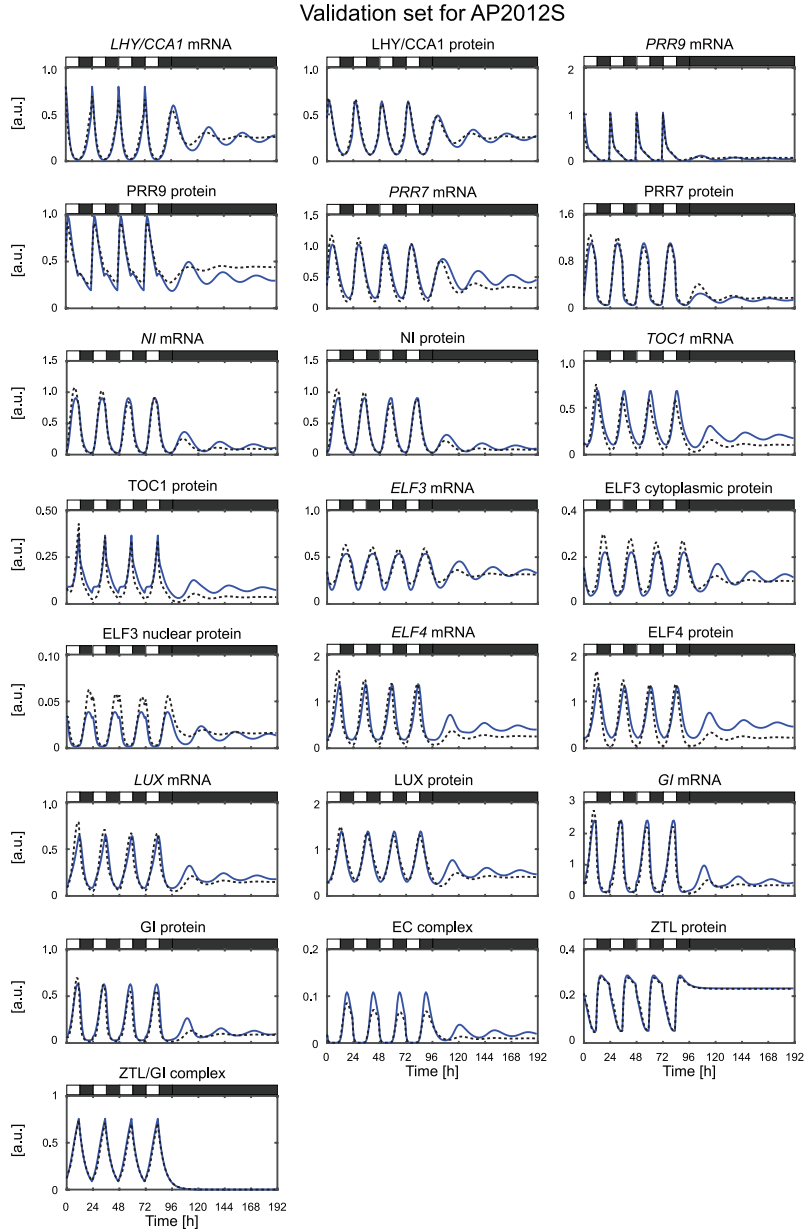
Validation set for JL2006S



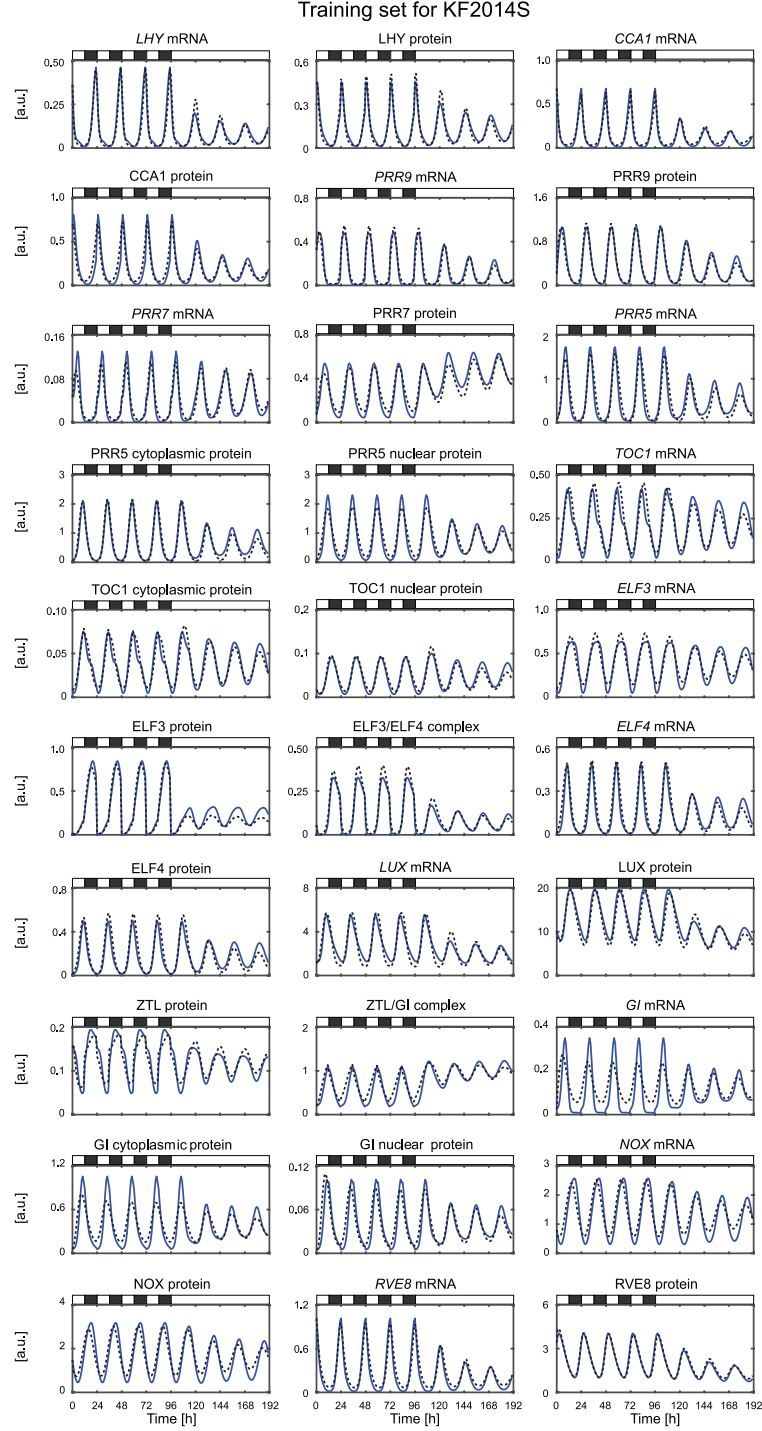
S2 Fig. Plant clock model JL2006S – fits to synthetic validation data. Blue solid lines show timeseries generated by JL2006 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant dark (DD). Black dashed lines show timeseries generated by JL2006S in the same simulated light environment using the parameters optimised to the training data (see **Fig 2C** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.



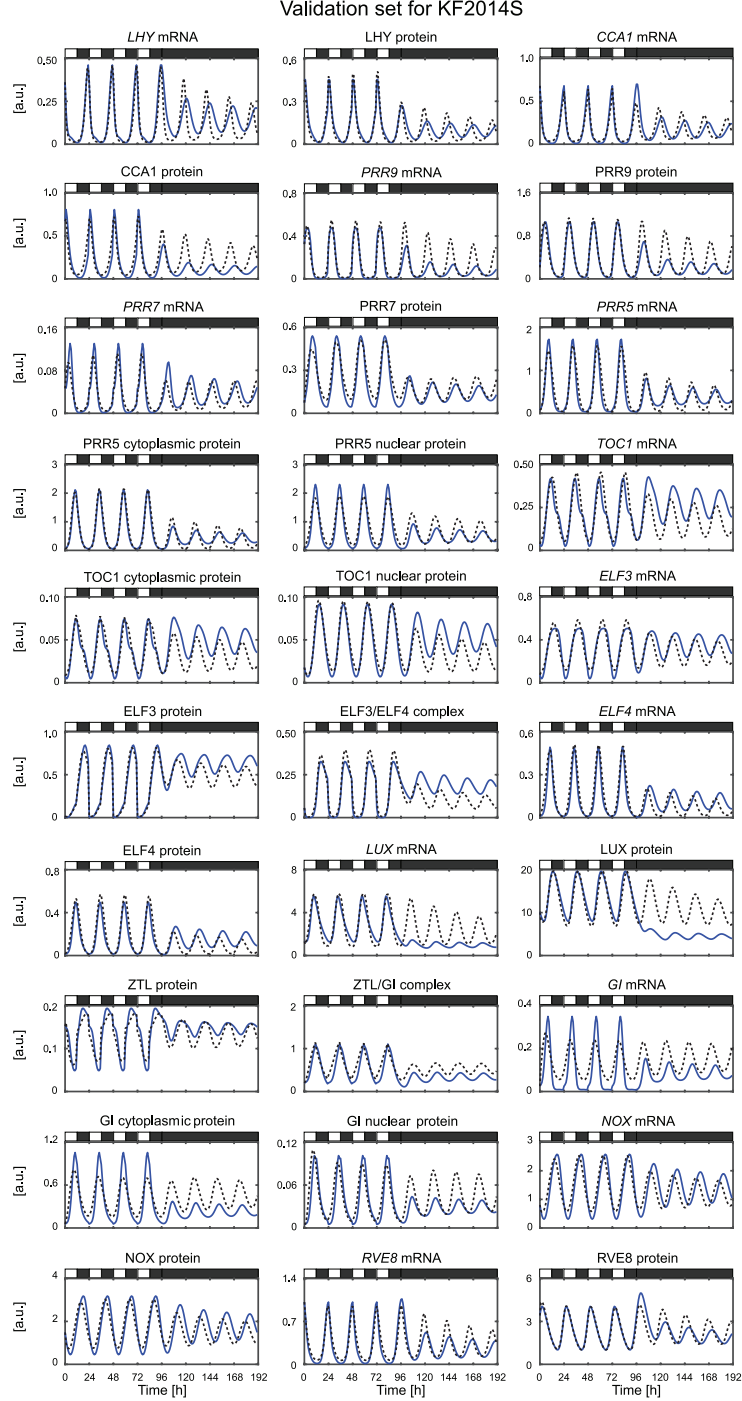
S3 Fig. Plant clock model AP2012S – optimal fits to synthetic training data. Blue solid lines show timeseries generated by AP2012 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant light (LL). Black dashed lines show timeseries obtained by optimising the parameters of AP2012S to this data in the same simulated light environment (see **Fig 3B** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.



S4 Fig. Plant clock model AP2012S – fits to synthetic validation data. Blue solid lines show timeseries generated by AP2012 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant dark (DD). Black dashed lines show timeseries generated by AP2012S in the same simulated light environment using the parameters optimised to the training data (see **Fig 3C** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

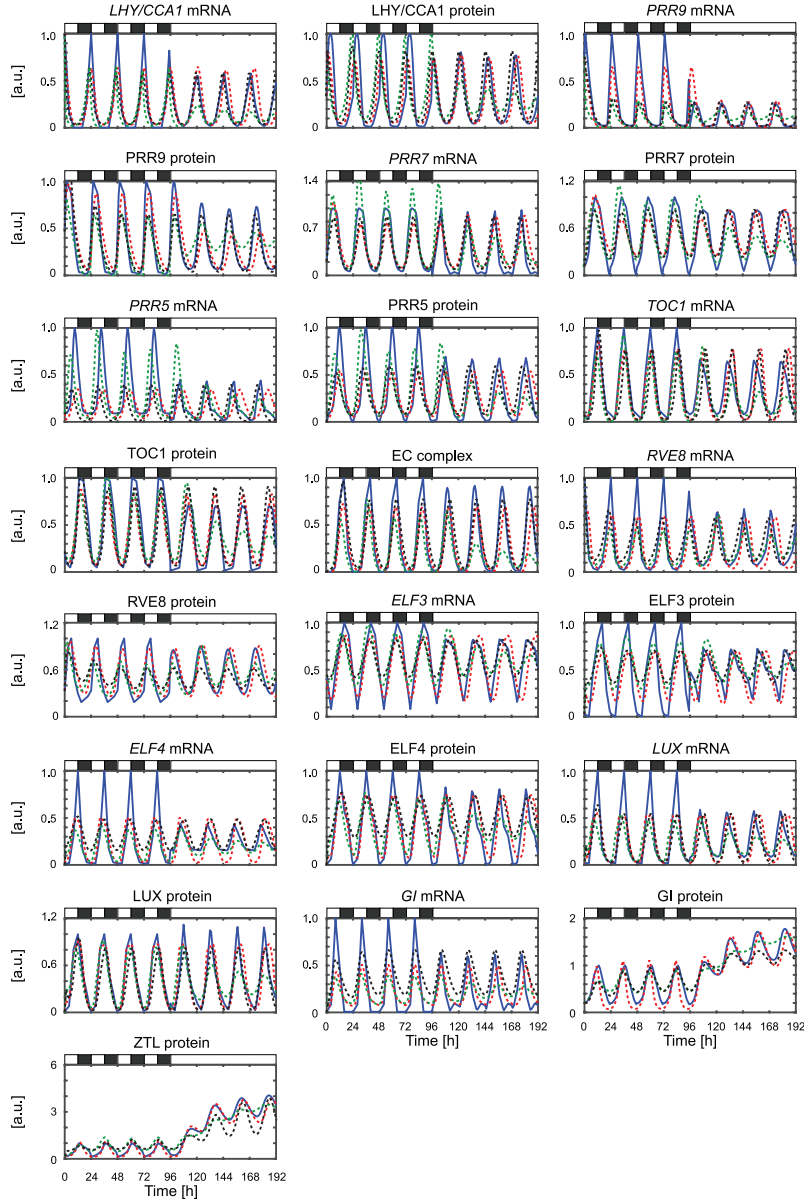


S5 Fig. Plant clock model KF2014S – optimal fits to synthetic training data. Blue solid lines show timeseries generated by KF2014 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant light (LL). Black dashed lines show timeseries obtained by optimising the parameters of KF2014S to this data in the same simulated light environment (see **Fig 4B** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

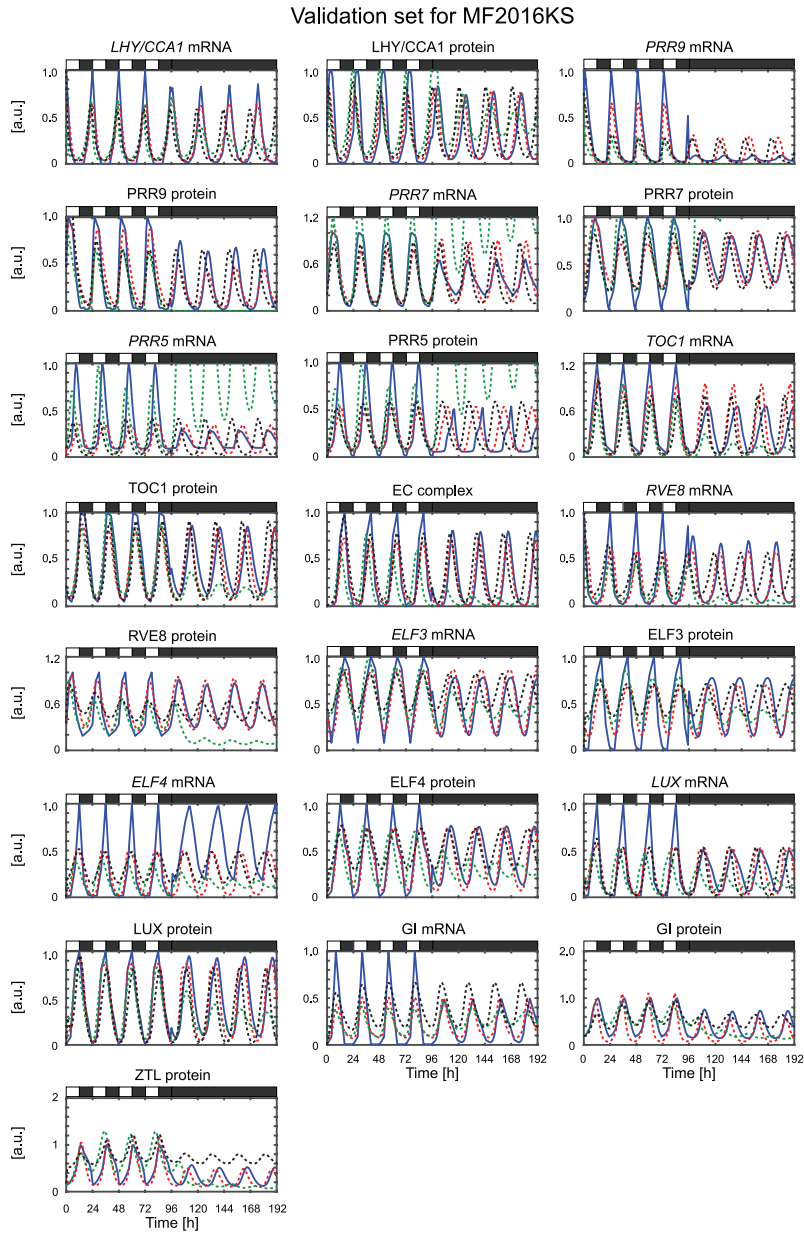


S6 Fig. Plant clock model KF2014S – fits to synthetic validation data. Blue solid lines show timeseries generated by KF2014 from its nominal parameter values for a simulated transition between a 12L:12D light-dark cycle and constant dark (DD). Black dashed lines show timeseries generated by KF2014S in the same simulated light environment using the parameters optimised to the training data (see **Fig 4C** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.

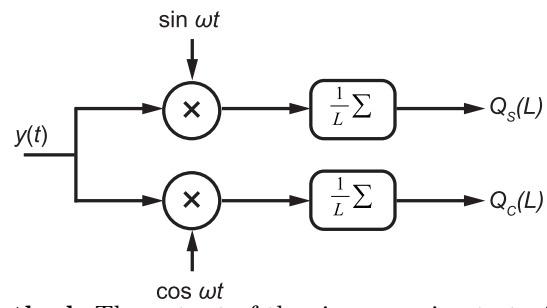
Training set for MF2016KS



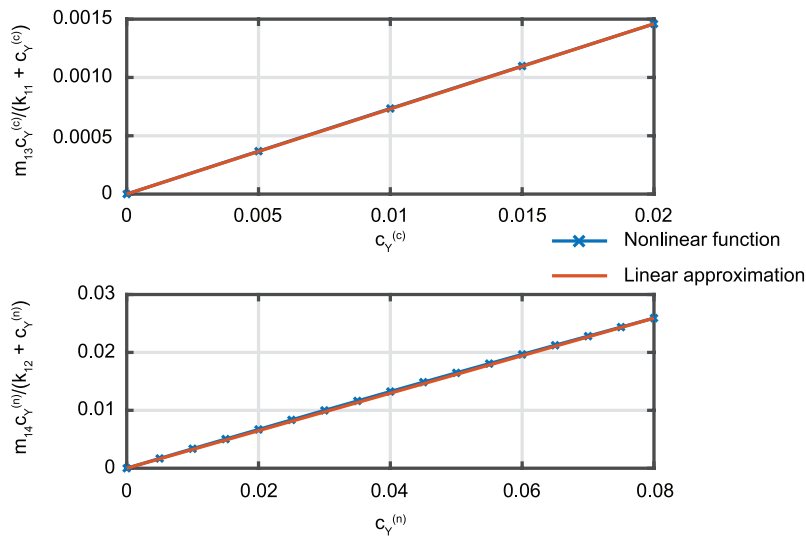
S7 Fig. Plant clock models MF2016KS, MF2016KSorig and MF2016K – optimal fits to experimental training data. Blue solid lines show timeseries recorded experimentally during a transition between a 12L:12D light-dark cycle and constant light (LL). Black (MF2016KS) and green (MF2016KSorig) dashed lines show timeseries obtained by optimising the parameters of the S-System models to this data in the same simulated light environment. Red dashed lines show optimal fits of MF2016K to the same data, obtained previously in [15] (see **Fig 5B** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.



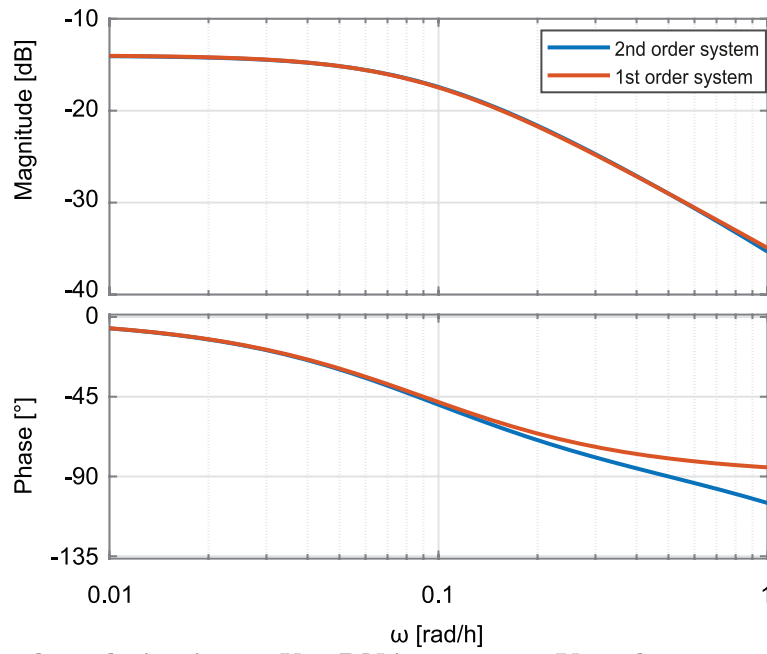
S8 Fig. Plant clock models MF2016KS, MF2016KSorig and MF2016K – fits to experimental validation data. Blue solid lines show timeseries recorded experimentally during a transition between a 12L:12D light-dark cycle and constant dark (DD). Black (MF2016KS) and green (MF2016KSorig) dashed lines show timeseries generated by the S-System models in the same simulated light environment using the parameters optimised to the training data. Red dashed lines show the corresponding fits of MF2016K to the same data (see **Fig 5C** for the corresponding heatmaps). White and black bars at the top of the figure indicate light and dark intervals, respectively.



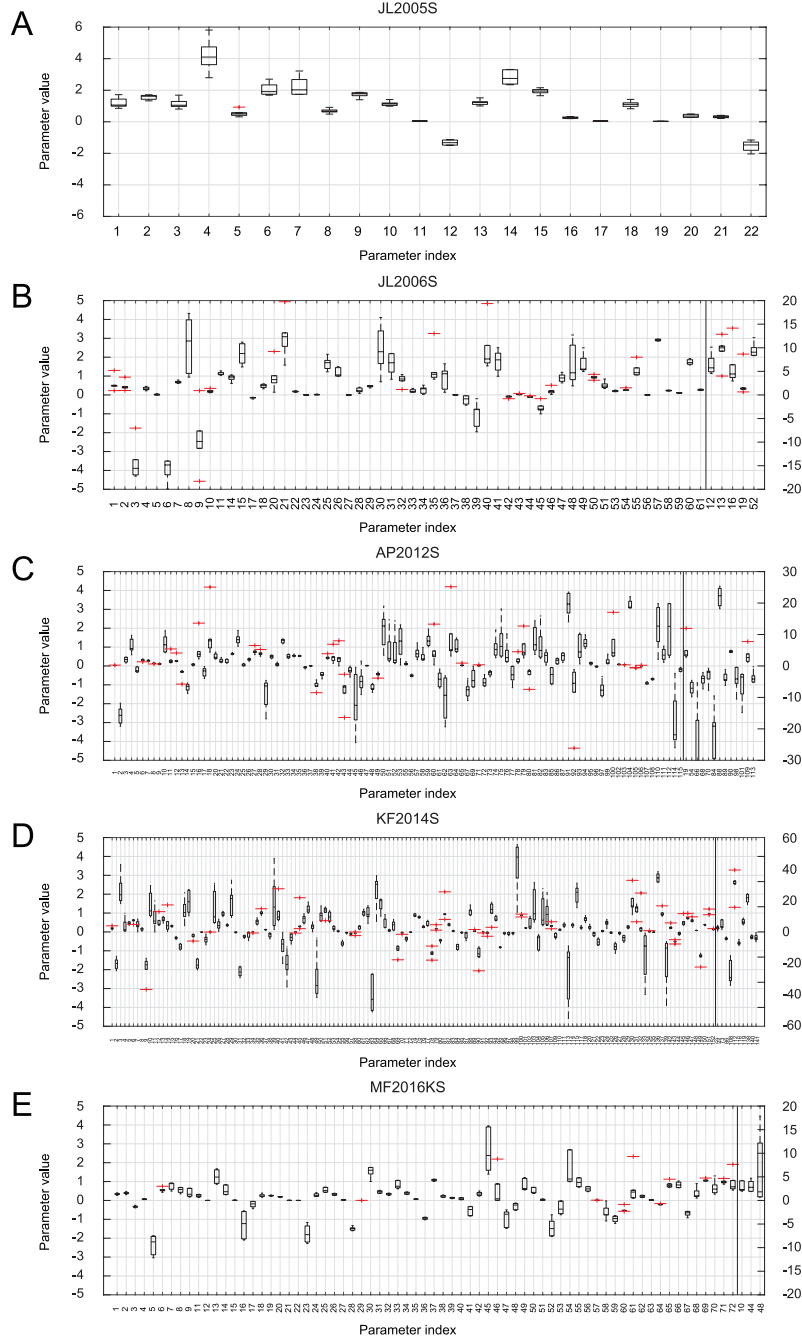
S9 Fig. Correlation method. The output of the sine sweeping test $y(t)$ is correlated with $\sin \omega t$ and $\cos \omega t$ prior to averaging to obtain the corresponding magnitude and phase values required to construct a Bode plot.



S10 Fig. Linear approximations to nonlinear Y protein degradation in JL2006. Blue lines show how the degradation rates of cytoplasmic Y protein (top panel) and nuclear Y protein (bottom panel) depend on the corresponding expression levels, $c_Y^{(c)}$ and $c_Y^{(n)}$, respectively. In each case, degradation rate is plotted for expression levels ranging between 0 and the maximum level observed in the synthetic training and validation datasets (see **S1 Fig** and **S2 Fig**). In these ranges, the nonlinear functions are well-approximated by linear fits (red lines), the gradients of which are taken as the values of γ_{cy} and γ_{nu} used to derive eq. (S3.5) in **S1 Text**.



S11 Fig. Bode plot relating input Y mRNA to output Y nuclear protein in JL2006. Blue lines represent the second-order system given by eq. (S3.5) in S1 Text. Red lines represent the first-order system given by eq. (S3.7) that approximates eq. (S3.5).



S12 Fig. Variation in optimised parameter values for the extended S-System models. **A-D:** Fits of JL2005S, JL2006S, AP2012S and KF2014S to synthetic data. **E:** Fits of MF2016KS to experimental data. Boxplots show parameter distributions obtained from six independent optimisation runs. In each boxplot, the horizontal line denotes the median value, the edges of the box are the 25th and 75th percentiles, the whiskers denote the most extreme datapoints not considered to be outliers, and outliers are plotted as red crosses. Model parameter indices are defined in **S3 Table** (JL2005S), **S6 Table** (JL2006S), **S9 Table** (AP2012S), **S12 Table** (KF2014S) and **S14 Table** (MF2016KS). In **B-E**, the thick black horizontal lines separate parameters whose values are plotted with respect to the left and right y-axes.

Supplementary Tables

i	Gene
L	$LHY/CCA1$ or LHY
T	$TOC1$
X	X
Y	Y/GI
A	$PRR7/9$
$P9$	$PRR9$
$P7$	$PRR7$
$NI/P5$	$PRR5$
$Lmod$	Modified $LHY/CCA1$ protein
$Ltot$	Total amount of $LHY/CCA1$ protein
$E4$	$ELF4$
LUX	LUX
ZTL	ZTL
$E3$	$ELF3$
GI	GI
$E3c$	$ELF3$ cytoplasmic protein
Gc	GI cytoplasmic protein
$COP1c$	$COP1$ cytoplasmic protein
$E3n$	$ELF3$ nuclear protein
Gn	GI nuclear protein
$COP1d$	$COP1$ nuclear protein – day form
$COP1n$	$COP1$ nuclear protein – night form
EGc	$ELF3$ - GI cytoplasmic protein complex
ZG	ZTL - GI cytoplasmic protein complex
EGn	$ELF3$ - GI nuclear protein complex
$E34$	$ELF3$ - $ELF4$ nuclear protein complex
EC	$ELF3$ - $ELF4$ - LUX nuclear protein complex
Gn_{tot}	Total amount of GI nuclear protein
$R/R8$	$RVE8$
C	$CCA1$
$E4d$	$ELF4$ dimer
NOX	NOX
$COP1$	$COP1$

S1 Table. Variables used in the equations for plant clock models JL2005 [11], JL2006 [43], AP2012 [44], KF2014 [45] and MF2016K [15].

Parameter	Value	Parameter	Value	Parameter	Value
a	1	b	2	n_1	16.9711
n_2	1.3043	g_1	3.0351	g_2	0.3860
m_1	9.3383	m_2	16.9058	m_3	1.0214
m_4	1.6859	m_5	0.4212	m_6	0.0484
m_7	1.2	k_1	1.3294	k_2	0.8085
k_3	0.1445	k_4	0.2089	k_5	0.3187
k_6	0.3505	k_7	1.2	p_1	4.9753
p_2	1.2947	p_3	0.5	r_1	1.4563
r_2	0.8421	r_3	0.0451	r_4	0.0018
q_1	2.9741	q_2	1	—	—

S2 Table. Nominal parameter values for JL2005, which were used to generate synthetic data. The parameters were taken from Figure 5 of [11].

Parameter	Value (nMAD)	Parameter	Value (nMAD)	Parameter	Value (nMAD)	
α_1	1.4361 (0.1320)	$g_{1,1}$	1.3276 (0.0439)	$\beta_{1,1}$	1.2875 (0.1499)	1–3
$\gamma_{1,1}$	4.7463 (0.1370)	α_2	0.4074 (0.1688)	$g_{2,1}$	2.0764 (0.1027)	4–6
$\beta_{2,1}$	1.7443 (0.1415)	α_3	0.9147 (0.0912)	$g_{3,1}$	1.7344 (0.0445)	7–9
$\beta_{3,1}$	1.4088 (0.0675)	α_4	0.0484 (0.1139)	$g_{4,1}$	−1.4874 (0.1198)	10–12
$\beta_{4,1}$	1.5163 (0.0687)	α_5	2.3917 (0.1373)	$g_{5,1}$	1.9657 (0.0519)	13–15
$\beta_{5,1}$	0.2075 (0.1665)	α_6	0.0378 (0.1310)	$g_{6,1}$	1.1309 (0.1009)	16–18
$\beta_{6,1}$	0.0268 (0.1353)	$\beta_{7,1}$	0.4886 (0.2460)	$\gamma_{7,1}$	0.4080 (0.1910)	19–21
$\gamma_{7,2}$	−1.5690 (0.1698)	—	—	—	—	22

S3 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation JL2005S of JL2005, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in **S12 Fig**.

Training dataset		Validation dataset	
Gene/protein	$W_i \left(X_i^{LL}, \hat{X}_i^{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-3})$	Gene/protein	$W_i \left(X_i^{DD}, \hat{X}_i^{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-3})$
<i>LHY/CCA1</i>	0.379	<i>LHY/CCA1</i>	0.388
LHY/CCA1 (cyt.)	0.393	LHY/CCA1 (cyt.)	0.405
LHY/CCA1 (nuc.)	1.005	LHY/CCA1 (nuc.)	0.998
<i>TOC1</i>	3.097	<i>TOC1</i>	2.915
TOC1 (cyt.)	4.461	TOC1 (cyt.)	4.530
TOC1 (nuc.)	0.179	TOC1 (nuc.)	0.196
Protein P	0.094	Protein P	0.407
$W \left(\mathbf{X}_{LL}, \hat{\mathbf{X}}_{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-3})$	1.372	$W \left(\mathbf{X}_{DD}, \hat{\mathbf{X}}_{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-3})$	1.406

S4 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting JL2005S to the synthetic training and validation datasets.

Parameter	Value	Parameter	Value	Parameter	Value
q_1	4.1954	n_0	0.0500	g_0	1.0000
α	4.0000	n_1	7.8142	a	1.2479
g_1	3.1383	m_1	1.9990	k_1	2.3920
p_1	0.8295	r_1	16.8363	r_2	0.1687
m_2	20.4400	k_2	1.5644	m_3	3.6888
k_3	1.2765	n_2	3.0087	b	1.0258
g_2	0.0368	g_3	0.2658	c	1.0258
m_4	3.8231	k_4	2.5734	p_2	4.3240
r_3	0.3166	r_4	2.1509	m_5	0.0013
m_6	3.1741	k_5	2.7454	m_7	0.0492
m_8	4.0424	k_6	0.4033	n_3	0.2431
d	1.4422	g_4	0.5388	m_9	10.1132
k_7	6.5585	p_3	2.1470	r_5	1.0352
r_6	3.3017	m_{10}	0.2179	k_8	0.6632
m_{11}	3.3442	k_9	17.1111	q_2	2.4017
n_4	0.0857	n_5	0.1649	g_5	1.1780
g_6	0.0645	e	3.6064	f	1.0237
m_{12}	4.2970	k_{10}	1.7303	p_4	0.2485
r_7	2.2123	r_8	0.2002	m_{13}	0.1347
k_{11}	1.8258	m_{14}	0.6114	k_{12}	1.8066
p_5	0.5000	k_{13}	1.2000	m_{15}	1.2000
q_3	1.0000	q_4	2.4514	g	1.0258
n_6	8.0706	g_7	0.0004	m_{16}	12.2398
k_{14}	10.3617	p_6	0.2907	r_9	0.2528
r_{10}	0.2212	m_{17}	4.4505	k_{15}	0.0703
m_{18}	0.0156	k_{16}	0.6104	—	—

S5 Table. Nominal parameter values for JL2006, which were used to generate synthetic data. The parameter values were taken from inline Supplementary Table 1 of [43].

Parameter	Value (nMAD)	Parameter	Value (nMAD)	Parameter	Value (nMAD)	
$\beta_{1,1}$	0.4597 (0.0503)	$\gamma_{1,1}$	0.3771 (0.0879)	$\gamma_{1,2}$	-4.2379 (0.0974)	1–3
α_2	0.4537 (0.1982)	$g_{2,1}$	0.0012 (0.7657)	$g_{2,2}$	-3.5046 (0.0566)	4–6
$\beta_{2,1}$	0.6117 (0.0567)	$\gamma_{2,1}$	3.0769 (0.4517)	$g_{2,3}$	-2.8305 (0.1872)	7–9
α_3	0.1151 (0.1971)	$g_{3,1}$	1.0993 (0.0323)	$\beta_{3,1}$	4.5548 (0.1768)	10–12
α_4	9.2696 (0.0549)	$g_{4,1}$	0.8282 (0.0967)	$\beta_{4,1}$	2.7978 (0.2342)	13–15
α_5	3.6836 (0.2521)	$g_{5,1}$	-0.2164 (0.0327)	$g_{5,2}$	0.5232 (0.1600)	16–18
$\beta_{5,1}$	1.2220 (0.0959)	α_6	2.3001 (0.2314)	$g_{6,1}$	1.5802 (0.1162)	19–21
$\beta_{6,1}$	0.2226 (0.1236)	$\gamma_{6,1}$	-0.0002 (0.2330)	α_7	0.0067 (0.6384)	22–24
$g_{7,1}$	1.8217 (0.1254)	$\beta_{7,1}$	1.0125 (0.0600)	$\gamma_{7,1}$	-0.0001 (0.7195)	25–27
α_8	0.1676 (0.4443)	$g_{8,1}$	0.4854 (0.0384)	$\beta_{8,1}$	1.7277 (0.3775)	28–30
α_9	2.1878 (0.2868)	$g_{9,1}$	1.0808 (0.1317)	$\beta_{9,1}$	0.1480 (0.1565)	31–33
α_{10}	0.3147 (0.3409)	$g_{10,1}$	1.1689 (0.1153)	$\beta_{10,1}$	1.1132 (0.2951)	34–36
α_{11}	0.0024 (0.6377)	$g_{11,1}$	-0.5448 (0.7978)	$g_{11,2}$	-1.6621 (0.3862)	37–39
$\beta_{11,1}$	1.8331 (0.1509)	$\gamma_{11,1}$	1.8447 (0.2400)	$g_{11,3}$	-0.0613 (0.2557)	40–42
$\gamma_{11,2}$	0.0150 (0.3048)	$g_{11,4}$	-0.1121 (0.0707)	$g_{11,5}$	-0.1966 (0.1436)	43–45
α_{12}	0.1160 (0.2427)	$g_{12,1}$	0.7780 (0.1837)	$\beta_{12,1}$	1.3022 (0.4499)	46–48
α_{13}	1.9436 (0.0615)	$g_{13,1}$	0.9501 (0.0406)	$\beta_{13,1}$	0.5854 (0.2023)	49–51
α_{14}	9.5387 (0.0763)	$g_{14,1}$	0.2558 (0.1048)	$\beta_{14,1}$	0.2512 (0.0113)	52–54
$\gamma_{14,1}$	1.0569 (0.0997)	α_{15}	0.0002 (0.0318)	$g_{15,1}$	2.9325 (0.0156)	55–57
$\beta_{15,1}$	0.2546 (0.0768)	α_{16}	0.1239 (0.0785)	$g_{16,1}$	1.6268 (0.0625)	58–60
$\beta_{16,1}$	0.2576 (0.0840)	—	—	—	—	61

S6 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation JL2006S of JL2006, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in **S12 Fig**.

Training dataset		Validation dataset	
Gene/protein	$W_i \left(X_i^{LL}, \hat{X}_i^{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	Gene/protein	$W_i \left(X_i^{DD}, \hat{X}_i^{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$
Protein P	0.076	Protein P	0.152
<i>LHY/CCA1</i>	0.356	<i>LHY/CCA1</i>	0.547
LHY/CCA1 (cyt.)	0.175	LHY/CCA1 (cyt.)	0.250
LHY/CCA1 (nuc.)	0.399	LHY/CCA1 (nuc.)	0.536
<i>TOC1</i>	0.549	<i>TOC1</i>	1.236
TOC1 (cyt.)	0.478	TOC1 (cyt.)	1.464
TOC1 (nuc.)	0.476	TOC1 (nuc.)	1.809
<i>X</i>	1.094	<i>X</i>	1.551
X (cyt.)	0.165	X (cyt.)	0.678
X (nuc.)	0.185	X (nuc.)	0.458
<i>Y</i>	0.929	<i>Y</i>	4.227
Y (cyt.)	0.552	Y (cyt.)	3.411
Y (nuc.)	0.539	Y (nuc.)	3.811
<i>PRR9/PRR7</i>	1.440	<i>PRR9/PRR7</i>	1.566
PRR9/PRR7 (cyt.)	3.294	PRR9/PRR7 (cyt.)	3.233
PRR9/PRR7 (nuc.)	1.171	PRR9/PRR7 (nuc.)	1.000
$W \left(\mathbf{X}_{LL}, \hat{\mathbf{X}}_{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	0.742	$W \left(\mathbf{X}_{DD}, \hat{\mathbf{X}}_{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	1.621

S7 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting JL2006S to the synthetic training and validation datasets.

Parameter	Value	Parameter	Value	Parameter	Value
n_1	2.600	n_2	0.640	n_3	0.290
n_4	0.070	n_5	0.230	n_6	20.000
n_7	0.200	n_8	0.500	n_9	0.200
n_{10}	0.400	n_{11}	0.600	n_{12}	12.500
n_{13}	1.300	n_{14}	0.100	m_1	0.540
m_2	0.240	m_3	0.200	m_4	0.200
m_5	0.300	m_6	0.300	m_7	0.700
m_8	0.400	m_9	1.100	m_{10}	1.000
m_{11}	1.000	m_{12}	1.000	m_{13}	0.320
m_{14}	0.400	m_{15}	0.700	m_{16}	0.500
m_{17}	0.500	m_{18}	3.400	m_{19}	0.200
m_{20}	0.600	m_{21}	0.080	m_{22}	0.100
m_{23}	1.800	m_{24}	0.100	m_{25}	1.800
m_{26}	0.500	m_{27}	0.100	m_{28}	20.000
m_{29}	5.000	m_{30}	3.000	m_{31}	0.300
m_{32}	0.200	m_{33}	13.000	m_{34}	0.600
m_{35}	0.300	m_{36}	0.100	m_{37}	0.800
m_{38}	0.500	m_{39}	0.300	p_1	0.130
p_2	0.270	p_3	0.100	p_4	0.560
p_5	4.000	p_6	0.600	p_7	0.300
p_8	0.600	p_9	0.800	p_{10}	0.540
p_{11}	0.510	p_{12}	3.400	p_{13}	0.100
p_{14}	0.140	p_{15}	3.000	p_{16}	0.620
p_{17}	4.800	p_{18}	4.000	p_{19}	1.000
p_{20}	0.100	p_{21}	1.000	p_{22}	0.500
p_{23}	0.370	p_{24}	10.000	p_{25}	8.000
p_{26}	0.300	p_{27}	0.800	p_{28}	2.000
p_{29}	0.100	p_{30}	0.900	p_{31}	0.100
g_1	0.100	g_2	0.010	g_3	0.600
g_4	0.010	g_5	0.150	g_6	0.300
g_7	0.600	g_8	0.010	g_9	0.300
g_{10}	0.500	g_{11}	0.700	g_{12}	0.200
g_{13}	1.000	g_{14}	0.004	g_{15}	0.400
g_{16}	0.300	a	2.000	b	2.000
c	2.000	d	2.000	e	2.000
f	2.000	q_1	1.000	q_2	1.560
q_3	2.800	—	—	—	—

S8 Table. Nominal parameter values for AP2012, used to generate synthetic data. The parameter values were taken from Supplemental Table 1 of [44].

Parameter	Value (nMAD)	Parameter	Value (nMAD)	Parameter	Value (nMAD)	
α_1	0.0290 (0.1534)	$g_{1,1}$	-1.9737 (0.1415)	$\beta_{1,1}$	0.2061 (0.3579)	1-3
$\gamma_{1,1}$	0.5920 (0.2203)	$\gamma_{1,2}$	-0.2176 (0.2879)	α_2	0.2947 (0.0596)	4-6
$\beta_{2,1}$	0.2925 (0.0727)	$\gamma_{2,1}$	0.1456 (0.0350)	α_3	0.1305 (0.1813)	7-9
$g_{3,1}$	1.8637 (0.3554)	$\beta_{3,1}$	0.1863 (0.2494)	$\gamma_{4,1}$	0.2521 (0.0768)	10-12
$\gamma_{4,2}$	-0.2893 (0.0921)	$\gamma_{4,3}$	-0.9038 (0.1275)	α_5	0.0559 (0.5552)	13-15
$g_{5,1}$	0.5870 (0.2165)	$g_{5,2}$	-0.2868 (0.6610)	$\beta_{5,1}$	1.3060 (0.1714)	16-18
$\gamma_{5,1}$	3.3540 (0.1516)	α_6	0.5235 (0.1730)	$\beta_{6,1}$	0.2691 (0.2261)	19-21
$\gamma_{6,1}$	0.1829 (0.1442)	α_7	0.7157 (0.0326)	$g_{7,1}$	1.5493 (0.1140)	22-24
$g_{7,2}$	0.0207 (0.9377)	$\beta_{7,1}$	0.3529 (0.1056)	α_8	0.7861 (0.0512)	25-27
$\beta_{8,1}$	0.6875 (0.0688)	$\gamma_{8,1}$	-0.7648 (0.2123)	α_9	0.5541 (0.1388)	28-30
$g_{9,1}$	0.0147 (0.8411)	$g_{9,2}$	1.3107 (0.0387)	$\beta_{9,1}$	0.6289 (0.1578)	31-33
α_{10}	0.5433 (0.0438)	$\beta_{10,1}$	0.5071 (0.0437)	$\gamma_{10,1}$	-0.0970 (0.2574)	34-36
α_{11}	0.0002 (0.4581)	$g_{11,1}$	-1.4163 (0.0776)	$g_{11,2}$	-0.6852 (0.1574)	37-39
$\beta_{11,1}$	0.3701 (0.0827)	α_{12}	0.4783 (0.1988)	$\beta_{12,1}$	0.4027 (0.2668)	40-42
$\beta_{12,2}$	1.1809 (0.1740)	$\beta_{12,3}$	0.3230 (0.4737)	$\gamma_{12,1}$	-2.8598 (0.5763)	43-45
$\gamma_{12,2}$	-0.7289 (0.3715)	α_{13}	0.0038 (0.3732)	$g_{13,1}$	-0.9857 (0.0849)	46-48
$g_{13,2}$	-0.6447 (0.0934)	$\beta_{13,1}$	1.1210 (0.3186)	α_{14}	0.3801 (0.3350)	49-51
$\beta_{14,1}$	0.2772 (0.4896)	$\beta_{14,2}$	-0.9820 (0.5148)	$\beta_{14,3}$	9.7979 (0.2392)	52-54
α_{15}	0.0708 (0.3394)	$g_{15,1}$	-0.5501 (0.0778)	$\beta_{15,1}$	0.4745 (0.2802)	55-57
α_{16}	0.6022 (0.1699)	$g_{16,1}$	1.5735 (0.1845)	$\beta_{16,1}$	0.7994 (0.2639)	58-60
$\beta_{16,2}$	0.3805 (0.5153)	$\beta_{16,3}$	0.8501 (0.6288)	α_{17}	0.8086 (0.2224)	61-63
$g_{17,1}$	0.8162 (0.1737)	$\beta_{17,1}$	0.0319 (0.5023)	$\beta_{17,2}$	27.0370 (0.3123)	64-66
$\beta_{17,3}$	1.3398 (0.1968)	$\beta_{17,4}$	3.2021 (0.2828)	$\beta_{17,5}$	7.7305 (0.4394)	67-69
$\beta_{17,6}$	1.1152 (0.4048)	α_{18}	0.0029 (0.8184)	$g_{18,1}$	-0.8698 (0.1971)	70-72
$g_{18,2}$	-0.5739 (0.2616)	$\beta_{18,1}$	0.8866 (0.3467)	α_{19}	0.5398 (0.5227)	73-75
$\beta_{19,1}$	0.1984 (0.5155)	$\beta_{19,2}$	0.7364 (0.7702)	α_{20}	0.2307 (0.2429)	76-78
$\beta_{20,1}$	0.7061 (0.2673)	$\gamma_{20,1}$	-0.2873 (0.2455)	α_{21}	0.6099 (0.3554)	79-81
$g_{21,1}$	0.9981 (0.4707)	$\beta_{21,1}$	0.1988 (0.3901)	$\gamma_{21,1}$	-19.7551 (0.2803)	82-84
$\gamma_{21,2}$	-0.2852 (0.8160)	α_{22}	0.0999 (0.3563)	$\beta_{22,1}$	0.3003 (0.2649)	85-87
$\gamma_{22,1}$	19.1516 (0.1202)	$\gamma_{22,2}$	-4.6314 (0.2319)	α_{23}	5.0022 (0.0584)	88-90
$\beta_{23,1}$	3.9124 (0.1589)	$\beta_{23,2}$	1.2039 (0.5806)	$\beta_{23,3}$	-0.1068 (0.6249)	91-93
α_{24}	1.1125 (0.1593)	$\beta_{24,1}$	0.1419 (0.3489)	$\beta_{24,2}$	0.0059 (0.5425)	94-96
$\beta_{24,3}$	1.0111 (0.2217)	$\gamma_{24,1}$	-5.5232 (0.6169)	α_{25}	0.1391 (0.2364)	97-99
$\beta_{25,1}$	0.6014 (0.2564)	$\gamma_{25,1}$	-3.7469 (0.2515)	$\gamma_{25,2}$	0.0920 (0.1545)	100-102
$\beta_{26,1}$	0.0805 (0.0431)	$\gamma_{26,1}$	3.4317 (0.0135)	$\gamma_{26,2}$	-0.0918 (0.0176)	103-105
α_{27}	0.0130 (0.2434)	$g_{27,1}$	-0.8984 (0.0410)	$g_{27,2}$	-0.7319 (0.0300)	106-108
$\beta_{27,1}$	3.8473 (0.4487)	$\gamma_{27,1}$	2.1840 (0.4260)	α_{28}	0.4535 (0.4570)	109-111
$\beta_{28,1}$	1.3219 (0.5762)	$\beta_{28,2}$	4.3258 (0.2449)	$\gamma_{28,1}$	-3.8148 (0.1300)	112-114
$\gamma_{28,2}$	-0.1205 (0.4523)	—	—	—	—	115

S9 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation AP2012S of AP2012, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in **S12 Fig**.

Training dataset		Validation dataset	
Gene/protein	$W_i \left(X_t^{LL}, \hat{X}_t^{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	Gene/protein	$W_i \left(X_t^{DD}, \hat{X}_t^{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$
<i>LHY/CCA1</i>	0.315	<i>LHY/CCA1</i>	0.361
<i>LHY/CCA1</i>	0.288	<i>LHY/CCA1</i>	0.186
<i>LHY/CCA1</i> (mod.)	0.244	<i>LHY/CCA1</i> (mod.)	0.994
Protein P	0.324	Protein P	1.113
<i>PRR9</i>	0.142	<i>PRR9</i>	0.140
<i>PRR9</i>	0.190	<i>PRR9</i>	0.697
<i>PRR7</i>	0.441	<i>PRR7</i>	1.119
<i>PRR7</i>	0.361	<i>PRR7</i>	0.358
<i>NI</i>	0.428	<i>NI</i>	0.470
<i>NI</i>	0.388	<i>NI</i>	0.392
<i>TOC1</i>	1.560	<i>TOC1</i>	1.877
<i>TOC1</i>	1.739	<i>TOC1</i>	1.486
<i>ELF3</i>	0.800	<i>ELF3</i>	0.727
<i>ELF3</i> (cyt.)	1.393	<i>ELF3</i> (cyt.)	1.654
<i>ELF3</i> (nuc.)	2.861	<i>ELF3</i> (nuc.)	7.345
<i>ELF4</i>	2.323	<i>ELF4</i>	1.811
<i>ELF4</i>	2.673	<i>ELF4</i>	3.643
<i>LUX</i>	1.573	<i>LUX</i>	1.082
<i>LUX</i>	1.833	<i>LUX</i>	0.777
<i>COP1</i> (cyt.)	0.004	<i>COP1</i> (cyt.)	0.003
<i>COP1</i> (nuc.)	0.031	<i>COP1</i> (nuc.)	0.031
<i>COP1</i> (dark)	0.031	<i>COP1</i> (dark)	0.029
<i>ELF3/GI</i> cplx.	5.575	<i>ELF3/GI</i> cplx.	1.058
<i>GI</i>	4.003	<i>GI</i>	0.720
<i>GI</i>	3.899	<i>GI</i>	0.986
<i>EC</i>	4.005	<i>EC</i>	1.906
<i>ZTL</i>	0.970	<i>ZTL</i>	0.074
<i>ZTL/GI</i> cplx.	0.703	<i>ZTL/GI</i> cplx.	0.148
$W \left(\mathbf{X}_{LL}, \hat{\mathbf{X}}_{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	1.396	$W \left(\mathbf{X}_{DD}, \hat{\mathbf{X}}_{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	1.114

S10 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting AP2012S to the synthetic training and validation datasets.

Parameter	Value	Parameter	Value	Parameter	Value
a_3	1.02200	a_4	9.86600	a_5	5.35300
a_6	2.27800	a_7	5.91700	a_8	4.36500
r_1	3.74700	r_2	2.38400	r_3	4.74700
r_4	16.30000	r_5	0.10000	r_6	0.47280
r_7	35.94000	r_8	2.22500	r_9	0.58850
r_{10}	21.80000	r_{11}	2.08600	r_{12}	6.03300
r_{13}	1.05300	r_{14}	12.66000	r_{15}	6.74300
r_{16}	0.15190	r_{17}	5.19900	r_{18}	1.20500
r_{19}	16.24000	r_{20}	0.14650	r_{21}	5.12700
r_{22}	1.97100	r_{23}	7.10000	r_{24}	16.33000
r_{25}	1.02700	r_{26}	5.46600	r_{27}	6.86400
r_{28}	8.39200	r_{29}	0.14230	r_{30}	2.71400
r_{31}	0.01041	r_{32}	4.77500	r_{33}	0.90260
r_{34}	0.05704	r_{35}	0.02929	r_{36}	0.49000
r_{37}	0.55400	r_{38}	0.05062	r_{40}	1.05100
r_{41}	0.33410	f_1	0.40900	f_2	2.02100
f_3	0.03313	f_4	0.10000	f_5	0.38530
f_6	0.25250	t_5	1.10300	t_6	0.58910
t_7	0.23170	t_8	0.14720	t_9	0.85430
m_1	0.99600	m_3	0.58890	m_4	0.37610
m_5	2.30000	m_6	0.01330	m_7	0.64870
m_8	5.43700	m_9	0.12250	m_{10}	0.01001
m_{11}	0.67710	m_{12}	1.98800	m_{13}	0.37600
m_{14}	4.91600	m_{15}	0.09303	m_{16}	0.58280
m_{17}	0.04744	m_{18}	2.42600	m_{19}	0.20000
m_{20}	1.80000	m_{21}	0.10000	m_{22}	0.30120
m_{23}	0.17640	m_{24}	2.84800	m_{25}	0.41760
m_{26}	1.57000	m_{27}	0.10000	m_{28}	0.02757
m_{29}	0.01000	m_{30}	5.49100	m_{31}	0.30000
m_{32}	5.68100	m_{33}	13.00000	m_{34}	0.11150
m_{35}	0.91880	m_{36}	0.57110	m_{37}	0.93910
m_{38}	7.97500	m_{39}	0.21600	m_{42}	0.37590
m_{43}	0.52140	m_{44}	0.35770	m_{45}	0.79440
m_{46}	7.54100	m_{47}	0.12930	p_6	0.60000
p_7	0.30000	p_{10}	0.20000	p_{11}	1.78000
p_{12}	8.00000	p_{13}	0.70000	p_{14}	0.30000
p_{15}	3.00000	p_{16}	0.40240	p_{23}	1.46100
p_{25}	1.11100	p_{28}	2.13000	p_{29}	25.20000
n_5	0.23000	n_6	20.00000	n_{14}	0.10000
q_1	0.12170	q_3	0.28730	—	—

S11 Table. Nominal parameter values for KF2014, which were used to generate synthetic data. The parameter values were taken from Table 3 (Parameter Set 2)[†] and Table 4 in Supporting Information Text S1 of [45] ([†]We note that in the original paper, Table 3 incorrectly lists parameter a_3 as a_1 – this has been fixed in our version of the table).

Parameter	Value (nMAD)	Parameter	Value (nMAD)	Parameter	Value (nMAD)	
α_1	0.1654 (0.1957)	$g_{1,1}$	-2.0558 (0.1338)	$\beta_{1,1}$	1.7392 (0.1980)	1-3
$\gamma_{1,1}$	0.1676 (0.6742)	α_2	0.4928 (0.1231)	$\beta_{2,1}$	0.6442 (0.0464)	4-6
$\gamma_{2,1}$	0.5449 (0.3852)	α_3	0.1344 (0.2636)	$g_{3,1}$	-1.5602 (0.1221)	7-9
$\beta_{3,1}$	0.8565 (0.2542)	$\gamma_{3,1}$	0.5030 (0.5635)	α_4	1.0750 (0.1336)	10-12
$\beta_{4,1}$	0.8219 (0.1316)	$\gamma_{4,1}$	0.0154 (0.6314)	$\gamma_{5,1}$	0.2759 (0.1062)	13-15
$\gamma_{5,2}$	-0.2823 (0.1282)	$\gamma_{5,3}$	-0.9646 (0.1703)	α_6	1.0913 (0.3099)	16-18
$g_{6,1}$	1.4395 (0.3645)	$g_{6,2}$	-0.1418 (0.2311)	$g_{6,3}$	-1.7397 (0.1179)	19-21
$g_{6,4}$	-0.0061 (0.5979)	$g_{6,5}$	-0.5398 (0.3225)	$g_{6,6}$	-0.00001 (0.4918)	22-24
$\beta_{6,1}$	0.8770 (0.4030)	$\gamma_{6,1}$	0.2028 (0.4994)	α_7	1.0909 (0.0882)	25-27
$\beta_{7,1}$	0.4303 (0.0934)	α_8	1.8273 (0.3023)	$g_{8,1}$	-0.0023 (0.8377)	28-30
$g_{8,2}$	-2.2365 (0.1103)	$g_{8,3}$	-0.1821 (0.2115)	$g_{8,4}$	-0.4600 (0.8290)	31-33
$g_{8,5}$	-0.0001 (0.6791)	$\beta_{8,1}$	0.4885 (0.1981)	α_9	1.0623 (0.0505)	34-36
$\beta_{9,1}$	0.1153 (0.0720)	$\gamma_{9,1}$	-0.0656 (0.5536)	α_{10}	0.3345 (0.7844)	37-39
$g_{10,1}$	1.0947 (0.2179)	$g_{10,2}$	-1.6729 (0.4925)	$g_{10,3}$	-1.2459 (0.2178)	40-42
$g_{10,4}$	-0.2541 (0.6014)	$g_{10,5}$	-0.0003 (0.5860)	$\beta_{10,1}$	1.8135 (0.0677)	43-45
α_{11}	0.9639 (0.2297)	$g_{11,1}$	1.0768 (0.1269)	$\beta_{11,1}$	0.5273 (0.3062)	46-48
$\beta_{11,2}$	1.9788 (0.1866)	α_{12}	1.3849 (0.2359)	$g_{12,1}$	0.5969 (0.0957)	49-51
$\beta_{12,1}$	1.1299 (0.3091)	α_{13}	0.2168 (0.4787)	$g_{13,1}$	0.0117 (0.6778)	52-54
$g_{13,2}$	-0.7642 (0.1369)	$g_{13,3}$	-0.0185 (0.4494)	$g_{13,4}$	-0.0537 (0.2395)	55-57
$g_{13,5}$	-0.0022 (0.2190)	$\beta_{13,1}$	4.0754 (0.2352)	α_{14}	0.1291 (0.1484)	58-60
$g_{14,1}$	0.8877 (0.0995)	$\beta_{14,1}$	0.4435 (0.2642)	$\beta_{14,2}$	3.3327 (0.1745)	61-63
α_{15}	2.9835 (0.1200)	$g_{15,1}$	1.7478 (0.0478)	$\beta_{15,1}$	0.2474 (0.0724)	64-66
α_{16}	0.0363 (0.3095)	$g_{16,1}$	0.2587 (0.5641)	$g_{16,2}$	-1.4757 (0.1274)	67-70
$g_{16,3}$	-0.0040 (0.3401)	$g_{16,4}$	-0.5084 (0.2028)	$g_{16,5}$	-0.0160 (0.7169)	70-72
$\beta_{16,1}$	9.8530 (0.3626)	α_{17}	0.9830 (0.0501)	$\beta_{17,1}$	0.8374 (0.0550)	73-75
α_{18}	0.2914 (0.1387)	$\beta_{18,1}$	0.3482 (0.0736)	$\beta_{18,2}$	1.4952 (0.0584)	76-78
α_{19}	0.0954 (0.0761)	$g_{19,1}$	-0.5924 (0.0846)	$\beta_{19,1}$	0.6617 (0.0188)	79-81
α_{20}	0.3773 (0.1210)	$\beta_{20,1}$	0.0911 (0.4273)	$\beta_{20,2}$	0.9687 (0.1628)	82-84
$\beta_{20,3}$	0.0102 (0.3749)	$\beta_{20,4}$	6.4433 (0.1107)	$\beta_{20,5}$	0.0077 (0.7329)	85-87
α_{21}	1.4329 (0.1150)	$\beta_{21,1}$	0.1281 (0.3959)	$\beta_{21,2}$	0.8576 (0.2013)	88-90
$\beta_{21,3}$	0.0123 (0.3396)	$\beta_{21,4}$	0.0019 (0.8196)	α_{22}	0.9879 (0.2061)	91-93
$g_{22,1}$	0.8199 (0.1257)	$g_{22,2}$	-0.8558 (0.0255)	$g_{22,3}$	-0.0400 (0.3991)	94-96
$g_{22,4}$	-0.0288 (0.7219)	$g_{22,5}$	-0.0170 (0.5535)	$\beta_{22,1}$	2.8521 (0.1502)	97-99
α_{23}	0.7927 (0.0158)	$\beta_{23,1}$	0.1775 (0.0482)	α_{24}	0.1696 (0.3730)	100-102
$\beta_{24,1}$	0.5155 (0.4003)	$\gamma_{24,1}$	-0.2182 (0.3942)	α_{25}	0.5644 (0.4300)	103-105
$g_{25,1}$	1.0340 (0.5346)	$\beta_{25,1}$	0.1672 (0.0666)	$\gamma_{25,1}$	-22.6595 (0.1185)	106-108
$\gamma_{25,2}$	-0.2877 (0.3616)	α_{26}	0.1043 (0.2607)	$\beta_{26,1}$	0.2457 (0.1609)	109-111
$\gamma_{26,1}$	19.5485 (0.0298)	$\gamma_{26,2}$	-4.5860 (0.3383)	α_{27}	0.3114 (0.1370)	112-114
$\beta_{27,1}$	1.5723 (0.1644)	$\beta_{27,2}$	8.0358 (0.0948)	$\gamma_{27,1}$	0.2021 (0.1401)	115-117
$\gamma_{27,2}$	0.6797 (0.1096)	α_{28}	7.5757 (0.2145)	$\beta_{28,1}$	0.1076 (0.2506)	118-120
$\gamma_{28,1}$	-0.2029 (0.7440)	$\gamma_{28,2}$	-0.7164 (0.2955)	α_{29}	0.0137 (0.2914)	121-123
$g_{29,1}$	0.4247 (0.1609)	$g_{29,2}$	-0.0043 (0.9354)	$g_{29,3}$	-0.6214 (0.2028)	124-126
$g_{29,4}$	-0.0041 (0.7542)	$g_{29,5}$	-0.5587 (0.3950)	$\beta_{29,1}$	0.2377 (0.2540)	127-129
α_{30}	1.7779 (0.1694)	$g_{30,1}$	1.4744 (0.1000)	$\beta_{30,1}$	0.2170 (0.4426)	130-132
$\beta_{30,2}$	0.1981 (0.8271)	$\gamma_{30,1}$	0.0026 (0.7952)	$\gamma_{30,2}$	-0.0002 (0.3149)	133-135
α_{31}	3.1990 (0.0760)	$g_{31,1}$	1.3817 (0.0602)	$\beta_{31,1}$	21.7247 (0.1037)	136-138
$\beta_{31,2}$	0.5155 (0.3443)	$\beta_{31,3}$	4.3588 (0.2647)	$\beta_{31,4}$	5.2823 (0.3421)	139-141
α_{32}	0.1731 (0.1577)	$g_{32,1}$	-0.6398 (0.0357)	$g_{32,2}$	-0.0134 (0.7423)	142-144
$\beta_{32,1}$	0.3799 (0.1749)	α_{33}	0.9841 (0.0441)	$\beta_{33,1}$	0.7988 (0.0527)	145-147
α_{34}	0.0913 (0.0449)	$g_{34,1}$	-1.8614 (0.0427)	$\beta_{34,1}$	0.3388 (0.0299)	148-150
α_{35}	1.2125 (0.0096)	$\beta_{35,1}$	0.1753 (0.0284)	—	—	151-152

S12 Table. Optimal parameter values $\hat{\Theta}_{LL}^S$ for the extended S-System formulation KF2014S of KF2014, obtained by fitting the model to the synthetic training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with those obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in **S12 Fig**.

Training dataset		Validation dataset	
Gene/protein	$W_i \left(\mathbf{X}_i^{LL}, \hat{\mathbf{X}}_i^{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	Gene/protein	$W_i \left(\mathbf{X}_i^{DD}, \hat{\mathbf{X}}_i^{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$
LHY	0.272	LHY	2.173
LHY	0.342	LHY	0.944
CCA1	0.315	CCA1	1.386
LHY	0.876	LHY	2.791
Protein P	0.141	Protein P	0.139
PRR9	0.426	PRR9	3.551
PRR9	0.260	PRR9	3.126
PRR7	0.851	PRR7	1.449
PRR7	0.790	PRR7	0.882
PRR5	0.899	PRR5	0.746
PRR5 (cyt.)	0.389	PRR5 (cyt.)	0.457
PRR5 (nuc.)	0.665	PRR5 (nuc.)	1.105
TOC1	1.311	TOC1	4.574
TOC1 (cyt.)	0.848	TOC1 (cyt.)	3.975
TOC1 (nuc.)	0.734	TOC1 (nuc.)	3.126
ELF3	0.957	ELF3	2.538
ELF3	0.446	ELF3	1.693
ELF3/4 cplx.	0.580	ELF3/4 cplx.	4.329
ELF4	0.797	ELF4	1.445
ELF4	1.315	ELF4	2.222
ELF4 (dark)	1.008	ELF4 (dark)	0.723
LUX	0.752	LUX	5.856
LUX	0.340	LUX	6.002
COP1 (cyt.)	0.013	COP1 (cyt.)	0.011
COP1 (nuc.)	0.051	COP1 (nuc.)	0.369
COP1 (dark)	0.098	COP1 (dark)	0.052
ZTL	0.678	ZTL	0.899
ZTL/GI cplx.	1.240	ZTL/GI cplx.	2.750
GI	2.841	GI	4.867
GI (cyt.)	1.565	GI (cyt.)	3.816
GI (nuc.)	0.014	GI (nuc.)	0.033
NOX	1.344	NOX	2.326
NOX	1.222	NOX	2.223
RVE8	0.525	RVE8	1.302
RVE8	0.102	RVE8	0.925
$W \left(\mathbf{X}_{LL}, \hat{\mathbf{X}}_{LL} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	0.757	$W \left(\mathbf{X}_{DD}, \hat{\mathbf{X}}_{DD} \left(\hat{\Theta}_{LL}^S \right) \right) (\times 10^{-2})$	2.240

S13 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting KF2014 to the synthetic training and validation datasets.

Parameter	Value (nMAD)	Parameter	Value (nMAD)	Parameter	Value (nMAD)	
α_1	0.2655 (0.0641)	$\beta_{1,1}$	0.3093 (0.0621)	$\gamma_{1,1}$	-0.2585 (0.0588)	1–3
α_2	0.0448 (0.2218)	$g_{2,1}$	-2.3152 (0.1407)	$\beta_{2,1}$	0.4314 (0.0594)	4–6
α_3	0.9112 (0.0275)	$\beta_{3,1}$	0.5660 (0.1865)	α_4	0.1780 (0.3857)	7–9
$g_{4,1}$	2.4323 (0.1074)	$\beta_{4,1}$	0.2501 (0.1580)	$\gamma_{4,1}$	0.00003 (0.8280)	10–12
α_5	1.5339 (0.2891)	$\beta_{5,1}$	0.5711 (0.3323)	α_6	0.0126 (0.4956)	13–15
$g_{6,1}$	-1.0164 (0.5353)	$g_{6,2}$	-0.2005 (0.7194)	$\beta_{6,1}$	0.3002 (0.1409)	16–18
α_7	0.2680 (0.0482)	$\beta_{7,1}$	0.2047 (0.0541)	α_8	0.0056 (0.3559)	19–21
$g_{8,1}$	-0.0002 (0.8012)	$g_{8,2}$	-1.3506 (0.2359)	$\beta_{8,1}$	0.4097 (0.1490)	22–24
α_9	0.5424 (0.1770)	$\beta_{9,1}$	0.2865 (0.0400)	α_{10}	0.0158 (0.4127)	25–27
$g_{10,1}$	-1.4978 (0.0413)	$g_{10,2}$	-0.0036 (0.0999)	$\beta_{10,1}$	1.0041 (0.0927)	28–30
α_{11}	0.5352 (0.1223)	$\beta_{11,1}$	0.3693 (0.0769)	α_{12}	0.7582 (0.1500)	31–33
$\beta_{12,1}$	0.3167 (0.1296)	α_{13}	0.0709 (0.0658)	$g_{13,1}$	-0.8926 (0.0513)	34–36
$\beta_{13,1}$	1.0452 (0.0182)	α_{14}	0.1568 (0.1125)	$\beta_{14,1}$	0.0930 (0.1129)	37–39
α_{15}	0.1438 (0.5179)	$g_{15,1}$	-0.3292 (0.3256)	$\beta_{15,1}$	0.3937 (0.2149)	40–42
α_{16}	3.1255 (0.3715)	$\beta_{16,1}$	3.5977 (0.3149)	α_{17}	2.1919 (0.9295)	43–45
$g_{17,1}$	-0.5002 (0.2369)	$g_{17,2}$	-0.1369 (0.3780)	$\beta_{17,1}$	17.8622 (0.5947)	46–48
α_{18}	0.6246 (0.1028)	$\beta_{18,1}$	0.3979 (0.1046)	α_{19}	0.0033 (0.8938)	49–51
$g_{19,1}$	-1.7339 (0.2632)	$g_{19,2}$	-0.7559 (0.5820)	$\beta_{19,1}$	1.0217 (0.1151)	52–54
α_{20}	1.1968 (0.2465)	$\beta_{20,1}$	0.7312 (0.1924)	α_{21}	0.0025 (0.3278)	55–57
$g_{21,1}$	-0.4076 (0.2263)	$g_{21,2}$	-1.2183 (0.1368)	$g_{21,3}$	-0.5346 (0.0086)	58–60
$\beta_{21,1}$	0.1365 (0.4582)	α_{22}	0.1667 (0.2345)	$\beta_{22,1}$	0.0462 (0.0550)	61–63
$\beta_{22,2}$	0.1660 (0.2186)	α_{23}	0.8759 (0.1023)	$\beta_{23,1}$	0.9016 (0.1738)	64–66
$\gamma_{23,1}$	-0.7762 (0.1125)	α_{24}	0.8941 (0.2909)	$\beta_{24,1}$	1.9106 (0.1944)	67–69
$\beta_{24,2}$	-1.2984 (0.0233)	$h_{24,1}$	1.0077 (0.3438)	$h_{24,2}$	1.0119 (0.0451)	70–72

S14 Table. Optimal parameter values $\hat{\Theta}_{LL}^E$ for the extended S-System formulation MF2016KS of MF2016K, obtained by fitting to the experimental training data. For each parameter, the number in brackets is the normalised median absolute deviation (nMAD). This is calculated using the value shown, together with the values obtained from five additional, independent optimisation runs. The rightmost column shows the parameter indexing, counting left to right across rows, that is used in **S12 Fig**.

Training dataset		Validation dataset	
Gene/protein	$W_i \left(D_i^{LL}, \hat{X}_i^{LL} \left(\hat{\Theta}_{LL}^E \right) \right) (\times 10^{-2})$	Gene/protein	$W_i \left(D_i^{DD}, \hat{X}_i^{DD} \left(\hat{\Theta}_{DD}^E \right) \right) (\times 10^{-2})$
Protein P	1.129	Protein P	1.398
<i>LHY/CCA1</i>	1.556	<i>LHY/CCA1</i>	7.479
<i>LHY/CCA1</i>	5.938	<i>LHY/CCA1</i>	15.590
<i>PRR9</i>	6.483	<i>PRR9</i>	75.888
<i>PRR9</i>	2.210	<i>PRR9</i>	10.053
<i>PRR7</i>	1.985	<i>PRR7</i>	7.371
<i>PRR7</i>	2.185	<i>PRR7</i>	5.475
<i>PRR5</i>	7.040	<i>PRR5</i>	23.256
<i>PRR5</i>	5.366	<i>PRR5</i>	20.200
<i>TOC1</i>	6.544	<i>TOC1</i>	16.076
<i>TOC1</i>	2.160	<i>TOC1</i>	7.310
<i>EC</i>	1.973	<i>EC</i>	7.654
<i>RVE8</i>	1.809	<i>RVE8</i>	9.331
<i>RVE8</i>	3.342	<i>RVE8</i>	5.313
<i>ELF3</i>	2.155	<i>ELF3</i>	8.870
<i>ELF3</i>	4.080	<i>ELF3</i>	8.904
<i>ELF4</i>	4.688	<i>ELF4</i>	9.684
<i>ELF4</i>	5.371	<i>ELF4</i>	7.174
<i>LUX</i>	3.084	<i>LUX</i>	10.656
<i>LUX</i>	1.821	<i>LUX</i>	7.245
<i>GI</i>	7.632	<i>GI</i>	18.080
<i>GI</i>	2.111	<i>GI</i>	5.779
<i>COP1</i>	0.703	<i>COP1</i>	0.444
<i>ZTL</i>	6.149	<i>ZTL</i>	36.542
$W \left(\mathbf{D}_{LL}, \hat{\mathbf{X}}_{LL} \left(\hat{\Theta}_{LL}^E \right) \right) (\times 10^{-2})$	3.647	$W \left(\mathbf{D}_{DD}, \hat{\mathbf{X}}_{DD} \left(\hat{\Theta}_{DD}^E \right) \right) (\times 10^{-2})$	13.557

S15 Table. The component-wise (W_i) and total (W) weighted mean squared error (WMSE) values obtained when fitting MF2016KS to the experimental training and validation datasets.

Parameter	Value	Parameter	Value	Parameter	Value
α_1	0.1233	$\beta_{1,1}$	0.8403	$h_{1,1}$	0.4190
$h_{1,2}$	0.6895	α_2	0.0588	$g_{2,1}$	-0.2158
$g_{2,2}$	-0.2705	$g_{2,3}$	-0.4991	$\beta_{2,1}$	1.9153
α_3	0.9834	$\beta_{3,1}$	0.4690	α_4	0.6552
$g_{4,1}$	0.7346	$g_{4,2}$	0.0309	$g_{4,3}$	1.3117
$\beta_{4,1}$	2.0981	α_5	1.3954	$\beta_{5,1}$	0.4393
α_6	0.0036	$g_{6,1}$	-2.2477	$g_{6,2}$	-0.0557
$\beta_{6,1}$	0.2495	α_7	0.2332	$\beta_{7,1}$	0.2131
α_8	0.0007	$g_{8,1}$	-0.00002	$g_{8,2}$	-2.9009
$\beta_{8,1}$	0.2543	α_9	0.5167	$\beta_{9,1}$	0.4078
α_{10}	0.0341	$g_{10,1}$	-1.6459	$g_{10,2}$	-0.0022
$\beta_{10,1}$	1.5880	α_{11}	0.3802	$\beta_{11,1}$	0.2508
α_{12}	0.9153	$\beta_{12,1}$	0.5608	α_{13}	0.0560
$g_{13,1}$	-0.8123	$\beta_{13,1}$	1.0326	α_{14}	0.3423
$\beta_{14,1}$	0.1800	α_{15}	0.0588	$g_{15,1}$	-0.7041
$\beta_{15,1}$	0.2594	α_{16}	1.8694	$\beta_{16,1}$	2.1363
α_{17}	0.0076	$g_{17,1}$	-1.5118	$g_{17,2}$	-0.4948
$\beta_{17,1}$	0.8314	α_{18}	1.2916	$\beta_{18,1}$	0.6880
α_{19}	0.0016	$g_{19,1}$	-2.1824	$g_{19,2}$	-0.8092
$\beta_{19,1}$	0.8623	α_{20}	1.1027	$\beta_{20,1}$	0.5859
α_{21}	0.0192	$g_{21,1}$	-0.7403	$g_{21,2}$	-0.5839
$g_{21,3}$	-0.4361	$\beta_{21,1}$	0.6222	α_{22}	0.1484
$\beta_{22,1}$	0.4972	$h_{22,1}$	2.1771	$h_{22,2}$	1.8459
$h_{22,3}$	1.3754	α_{23}	0.1013	$\beta_{23,1}$	1.2943
$h_{23,1}$	1.5038	$h_{23,2}$	1.3948	α_{24}	1.9545
$g_{24,1}$	0.0127	$g_{24,2}$	1.6218	$\beta_{24,1}$	1.2789

S16 Table. Optimal parameter values $\hat{\Theta}_{LL}^E$ for the original S-System formulation MF2016KSorig of MF2016K, obtained by fitting to the experimental training data.

Parameter	Value	Parameter	Value	Parameter	Value
ϕ_1	-0.4330	ϕ_2	-0.4949	ϕ_3	0.4358
ϕ_4	1.9601	ϕ_5	0.7412	ϕ_6	0.3821
ϕ_7	0.5264	ϕ_8	1.3764	ϕ_9	11.6684
ϕ_{10}	-2.7804	ϕ_{11}	0.8360	ϕ_{12}	-0.6472
ϕ_{13}	0.7337	ϕ_{14}	0.00006	ϕ_{15}	0.7598
ϕ_{16}	1.3352	ϕ_{17}	-0.5971	ϕ_{18}	0.5008
ϕ_{19}	-0.2234	ϕ_{20}	0.0099	ϕ_{21}	0.0347
ϕ_{22}	0.8773	ϕ_{23}	-0.2393	ϕ_{24}	0.2374
ϕ_{25}	-0.1857	ϕ_{26}	0.0267	ϕ_{27}	0.1320
ϕ_{28}	0.03159	ϕ_{29}	-40.9795	ϕ_{30}	4.0753
ϕ_{31}	-2.5872	ϕ_{32}	0.0150	ϕ_{33}	0.0369
ϕ_{34}	0.0113	ϕ_{35}	-12.2192	ϕ_{36}	0.4005
ϕ_{37}	-0.3117	ϕ_{38}	0.8082	ϕ_{39}	-0.3785
ϕ_{40}	2.2417	ϕ_{41}	1.4074	ϕ_{42}	-1.5159
ϕ_{43}	0.3285	ϕ_{44}	-0.1488	ϕ_{45}	0.0146
ϕ_{46}	0.2339	ϕ_{47}	-0.2726	ϕ_{48}	14.0123
ϕ_{49}	-17.3569	ϕ_{50}	0.0144	ϕ_{51}	0.0321
ϕ_{52}	0.0517	ϕ_{53}	-11.0070	ϕ_{54}	0.6736
ϕ_{55}	-0.3972	ϕ_{56}	0.0289	ϕ_{57}	0.0173
ϕ_{58}	0.0341	ϕ_{59}	-26.4166	ϕ_{60}	2.4610
ϕ_{61}	-1.4794	ϕ_{62}	0.0002	ϕ_{63}	0.0724
ϕ_{64}	0.1181	ϕ_{65}	0.4560	ϕ_{66}	-0.5845
ϕ_{67}	0.4025	ϕ_{68}	-0.7857	ϕ_{69}	-0.0208
ϕ_{70}	-0.8094	ϕ_{71}	-0.8102	ϕ_{72}	0.9142
ϕ_{73}	2.3354	ϕ_{74}	-0.4454	ϕ_{75}	0.0279
θ_{144}	1.275	θ_{145}	1.5983	—	—

S17 Table. Optimal parameter values for MF2016K, which were obtained previously in [15] by fitting to the experimental training data used in this study with the same optimisation method. The values are reproduced from Tables S2 and S4 in the Supporting Information of [15].